

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

GM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:35:01 ; Search time 58.67 Seconds

(without alignments)  
8658.089 Million cell updates/sec

Title: US-09-932-678-1

Sequence: 1 accagagcttggctggaag.....tgcatttggattcccat 2068

Scoring table: JDEENTITY\_MGC

Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents, NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTBUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48.2	2.3	8797	2	US-08-723-306-6
2	48.2	2.3	8797	5	PCT-US96-10041-6
3	48.2	2.3	11093	2	US-08-723-306-5
4	48.2	2.3	11093	5	PCT-US96-10041-5
5	40.4	2.0	7218	1	US-08-232-463-14
6	39.2	1.9	4937	1	US-08-038-682-3
7	39.2	1.9	4937	1	US-08-302-832-3
8	39.2	1.9	4937	2	US-08-530-198-3
9	39.2	1.9	4937	2	US-08-469-880-3
10	39.2	1.9	4937	2	US-08-728-470-3
11	39.2	1.9	4937	2	US-08-617-697-3
12	39.2	1.9	4937	4	US-08-719-641-3
13	35.2	1.7	5538	2	US-08-231-193A-55
14	35.2	1.7	5538	2	US-08-486-273A-55
15	35.2	1.7	5538	3	US-08-940-086A-55
16	35.2	1.7	5538	4	US-08-940-035A-55
17	34.6	1.7	2518	4	US-09-433-699-3
18	34.4	1.7	289	4	US-09-007-005-17
19	34.4	1.7	289	4	US-09-244-796-17
20	34.4	1.7	1307	4	US-09-282-305-17
21	34.4	1.7	9333	1	US-08-038-682-6
22	34.4	1.7	9333	1	US-08-302-832-6
23	34.4	1.7	9333	2	US-08-530-198-6
24	34.4	1.7	9333	2	US-08-469-880-6
25	34.4	1.7	9333	2	US-08-728-470-6
26	34.4	1.7	9333	2	US-08-617-697-6
27	34.4	1.7	9333	4	US-08-719-641-6

## ALIGNMENTS

28	44	1.6	1245	1	US-08-385-186-14	Sequence 14, Appl
C 29	33.4	1.6	10614	1	US-08-135-511-35	Sequence 35, Appl
C 30	33.4	1.6	10614	1	US-08-187-453-35	Sequence 35, Appl
C 31	33.2	1.6	1545	1	US-07-872-673B-2	Sequence 2, Appl
C 32	33.2	1.6	2454	1	US-07-872-673B-2	Sequence 2, Appl
C 33	32.6	1.6	19182	2	US-08-850-880-11	Sequence 11, Appl
C 34	32.6	1.6	19182	2	US-08-944-916-11	Sequence 11, Appl
C 35	32.4	1.6	2791	4	US-09-570-367-1	Sequence 1, Appl
C 36	32.2	1.6	4856	4	US-04-045-460-1	Sequence 1, Appl
C 37	32.2	1.6	12225	2	US-08-822-445-11	Sequence 11, Appl
C 38	32.2	1.6	12225	4	US-09-396-540-11	Sequence 11, Appl
C 39	32.2	1.6	12616	2	US-08-822-445-9	Sequence 9, Appl
C 40	32.2	1.6	12616	4	US-09-396-540-9	Sequence 9, Appl
C 41	32.2	1.6	24674	4	US-09-773-816-1	Sequence 1, Appl
C 42	32	1.5	1202	3	US-09-058-489-29	Sequence 29, Appl
C 43	32	1.5	1248	4	US-08-910-505-4	Sequence 4, Appl
C 44	32	1.5	1251	4	US-08-910-505-1	Sequence 1, Appl
C 45	31.8	1.5	1338	4	US-09-027-064-3	Sequence 3, Appl

RESULT 1  
US-08-723-306-6  
Sequence 6, Application US/08723306  
Patent No. 5856178  
GENERAL INFORMATION:  
APPLICANT: White PhD, Kenneth  
INVENTOR: White PhD, Kenneth  
APPLICANT: Reed, William  
TITLE OF INVENTION: Cassette for Expression of Lytic  
NUMBER OF INVENTIONS: Peptides in Mammalian Transgenic Organisms  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask Hill and Kossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City  
STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatScan Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/723,306  
FILING DATE:  
CLASSIFICATION: B00  
ALL-KEY/AGENT INFORMATION:  
NAME: Sweigert PhD, Susan R  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2549  
TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8797 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
FUNCTION: not relevant  
MULTIPLE TYPE: other nucleic acid  
DESCRIPTION: /dose - "Construct comprising pos  
DESCRIPTION: taupus beta casein 5' regulatory region plus genes encoding  
DESCRIPTION: amphipathic peptide and green fluorescent protein"  
HYPOTHEICAL: YES  
ANTI-SENSE: NO  
US-08-7-3-306-6  
Query Match 2.3% Score 48.2; DB 2; Length 8797;

[illegible]

RESULT 2  
 PCT 0036, 10041 %  
 SEQUENCE 6; APPLICATION PW/TUS9610041  
 GENERAL INFORMATION:  
 APPLICANT: White Pub, Kenneth  
 APPLICANT: Morley Pub, John  
 APPLICANT: Reed, William  
 TITLE OF INVENTION: Cassette for Expression of Lytic  
 TITLE OF INVENTION: Peptides for Mammalian Transgenic Organisms  
 NUMBER OF SEQUENCES: 82  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Task Cellt and Reesed  
 STREET: P.O. Box 2550  
 CITY: Salt Lake City  
 STATE: Utah  
 COUNTRY: USA  
 ZIP: 84110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC DOS/MKS DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.40  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PW/TUS9610041  
 FILING DATE:  
 CLASSIFICATION:  
 ALTERNATE/ACCENT INFORMATION:  
 NAME: Scott-Patt PhD, Susan E  
 REGISTRATION NUMBER: 66,209  
 REFERENCE/JOURNAL NUMBER: 2949  
 FIELD/COMMUNICATION INFORMATION:  
 TELEPHONE: 801/527922  
 TELEFAX: 801/5419148  
 INFORMATION FOR SEQ ID NO: 6;  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 8797 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 ORIENTATION: not relevant  
 MODIFIABLE TYPE: other nucleic acid  
 DESCRIPTION: Mouse beta casein 5' regulatory region plus genes encoding  
 DISCUSSION: amphipathic peptide and green fluorescent protein"  
 HYDROPHOBIC: YES  
 ANTI SENSE: NO  
 PCT 0036, 10041 %  
 GENEY MATCH: 2 3% SCORE 48.2; DB 5; Length 8797;  
 Most Local Similarity 92.2%; Prev. No. 0.00974;  
 Matches 107; Conservative 0; Mismatches 96; Indels 0; Caps 0;  
 597-

[illegible]

```

1  RESCUE 1 4 006 5
2  US-08 724-406 5
3  Sequence 5: Application US/0872406
4  Patient No. 5856178
5
6  GENERAL INFORMATION:
7  APPLICANT: White Phil, Kenneth
8  APPLICANT: Morrey Phil, John
9  APPLICANT: Reed, William
10 TITLE OF INVENTION: Cassette for Expression of Tyro
11 TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms
12 NUMBER OF SEQUENCES: 42
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: Truck Nitte and Kossak
15 STREET: P.O. Box 2750
16 CITY: Salt Lake City
17 STATE: Utah
18 COUNTRY: USA
19 ZIP: 84110
20
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC DOS/MS DOS
25 SOFTWARE: Patent In Release #1.0, Version #1.0
26
27 CURRENT APPLICATION DATA:
28 APPLICATION NUMBER: US/087241406
29
30 FILING DATE:
31 CLASSIFICATION: 600
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Societech, Inc.
34 REGISTRATION NUMBER: 46,289
35 REFERENCE/DOCKET NUMBER: 2349
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: 8015421922
38 TELEFAX: 8015419168
39 INFORMATION FOR SEQ ID NO: 5:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 1193 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: double
44 TECHNOLOGY: not relevant
45
46 MOLECULE TYPE: other nucleic acid
47 DESCRIPTION: /desc "Construct comprising
48 DESCRIPTION: /desc portions of post-taurus beta casein gene and genes encoding
49 DESCRIPTION: /desc amphipathic peptide and green fluorescent protein."
50 HYDROPHOBIC: YES
51 ANTI-SENSE: NO
52
53 FEATURE:
54 NAME/KEY: exon
55 LOCATION: 1801..1844
56 OTHER INFORMATION: /product "beta casein exon 1"
57
58 FEATURE:
59 NAME/KEY: exon
60 LOCATION: 4780..4842
61 OTHER INFORMATION: /product "beta casein exon 2"
62
63 FEATURE:
64 NAME/KEY: LATA_Signal
65 LOCATION: 1766..1774

```











```

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: US 08/302,812
3 FILING DATE: 16-SEP-1994
4
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: US PCT/US94/02166
7 FILING DATE: 16-MAR-1993
8
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: GB 9205704.1
11 FILING DATE: 16-MAR-1992
12
13 ATTORNEY/AGENT INFORMATION:
14 NAME: BOCKSTRESSER, JERRY W
15 REGISTRATION NUMBER: 22,651
16 REFERENCE/DOCKET NUMBER: 10388-625
17
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (703) 415-0810
20
21 TELEFAX: (703) 415-0814
22
23 INFORMATION FOR SEQ ID NO: 3:
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 4947 base pairs
26 TYPE: nucleic acid
27 STRANDEDNESS: single
28
29 1 topology: linear
30
31 MOLECULE TYPE: DNA (genomic)
32
33 OS-08-719-641-3
34
35 Query Match 1 98; Score 39.2; DB 4; Length 4947;
36 Host Local Similarity 53.28; Prod. No. 0.14;
37 Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

	Query Match	19%	Score 39.47	DB 4:	Length 4477
	Host Local Similarity	53.28%	Freqd. No.	0.14:	
Matches	83:	Conservative	0:	Mismatches	73:
				Indels	0:
				Gaps	0:
DY	BC3	tattatattaaacttaccatactgaagttgagaatggacccgcagggttatggaca	66%Z		
Db	4887	tgattatttaa-ccaacaaaatatcctttaaaccttaatttcgatcatcattcacctcatactaatca-tacacacat	44%Z		

Accession	Gene	Accession	Gene
U00096	16S rRNA	U00096	16S rRNA
U00097	23S rRNA	U00097	23S rRNA
U00098	5S rRNA	U00098	5S rRNA
U00099	16S rRNA	U00099	16S rRNA
U00100	23S rRNA	U00100	23S rRNA
U00101	5S rRNA	U00101	5S rRNA
U00102	16S rRNA	U00102	16S rRNA
U00103	23S rRNA	U00103	23S rRNA
U00104	5S rRNA	U00104	5S rRNA
U00105	16S rRNA	U00105	16S rRNA
U00106	23S rRNA	U00106	23S rRNA
U00107	5S rRNA	U00107	5S rRNA
U00108	16S rRNA	U00108	16S rRNA
U00109	23S rRNA	U00109	23S rRNA
U00110	5S rRNA	U00110	5S rRNA
U00111	16S rRNA	U00111	16S rRNA
U00112	23S rRNA	U00112	23S rRNA
U00113	5S rRNA	U00113	5S rRNA
U00114	16S rRNA	U00114	16S rRNA
U00115	23S rRNA	U00115	23S rRNA
U00116	5S rRNA	U00116	5S rRNA
U00117	16S rRNA	U00117	16S rRNA
U00118	23S rRNA	U00118	23S rRNA
U00119	5S rRNA	U00119	5S rRNA
U00120	16S rRNA	U00120	16S rRNA
U00121	23S rRNA	U00121	23S rRNA
U00122	5S rRNA	U00122	5S rRNA
U00123	16S rRNA	U00123	16S rRNA
U00124	23S rRNA	U00124	23S rRNA
U00125	5S rRNA	U00125	5S rRNA
U00126	16S rRNA	U00126	16S rRNA
U00127	23S rRNA	U00127	23S rRNA
U00128	5S rRNA	U00128	5S rRNA
U00129	16S rRNA	U00129	16S rRNA
U00130	23S rRNA	U00130	23S rRNA
U00131	5S rRNA	U00131	5S rRNA
U00132	16S rRNA	U00132	16S rRNA
U00133	23S rRNA	U00133	23S rRNA
U00134	5S rRNA	U00134	5S rRNA
U00135	16S rRNA	U00135	16S rRNA
U00136	23S rRNA	U00136	23S rRNA
U00137	5S rRNA	U00137	5S rRNA
U00138	16S rRNA	U00138	16S rRNA
U00139	23S rRNA	U00139	23S rRNA
U00140	5S rRNA	U00140	5S rRNA
U00141	16S rRNA	U00141	16S rRNA
U00142	23S rRNA	U00142	23S rRNA
U00143	5S rRNA	U00143	5S rRNA
U00144	16S rRNA	U00144	16S rRNA
U00145	23S rRNA	U00145	23S rRNA
U00146	5S rRNA	U00146	5S rRNA
U00147	16S rRNA	U00147	16S rRNA
U00148	23S rRNA	U00148	23S rRNA
U00149	5S rRNA	U00149	5S rRNA
U00150	16S rRNA	U00150	16S rRNA
U00151	23S rRNA	U00151	23S rRNA
U00152	5S rRNA	U00152	5S rRNA
U00153	16S rRNA	U00153	16S rRNA
U00154	23S rRNA	U00154	23S rRNA
U00155	5S rRNA	U00155	5S rRNA
U00156	16S rRNA	U00156	16S rRNA
U00157	23S rRNA	U00157	23S rRNA
U00158	5S rRNA	U00158	5S rRNA
U00159	16S rRNA	U00159	16S rRNA
U00160	23S rRNA	U00160	23S rRNA
U00161	5S rRNA	U00161	5S rRNA
U00162	16S rRNA	U00162	16S rRNA
U00163	23S rRNA	U00163	23S rRNA
U00164	5S rRNA	U00164	5S rRNA
U00165	16S rRNA	U00165	16S rRNA
U00166	23S rRNA	U00166	23S rRNA
U00167	5S rRNA	U00167	5S rRNA
U00168	16S rRNA	U00168	16S rRNA
U00169	23S rRNA	U00169	23S rRNA
U00170	5S rRNA	U00170	5S rRNA
U00171	16S rRNA	U00171	16S rRNA
U00172	23S rRNA	U00172	23S rRNA
U00173</			

4447 TATCACTGGGCTATTTAAACACAGTAATAGGTTAA 4442

RESULT : 3  
OS-08-431-193A-55

```

; GENERAL INFORMATION:
;
; APPLICANT: Daquet, Lorraine

```

1 AFFILIANT: Ellis, Steven B.  
2 AFFILIANT: Law, Chien W.  
3 AFFILIANT: Lu, Chin-Chou  
4 TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
5 TITLE OF INVENTION: SUBSTITUTED NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
6 NUMBER OF SEQUENCES: 63  
7 CORRESPONDENCE ADDRESS:  
8 ADDRESSEE: Brown, Martin, Haller & McLean

STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS DOS  
ETWARR: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/231,195A  
APPLICATION NUMBER: 536  
FILING DATE: 20-Apr-1994  
CLASSIFICATION: 536  
PRE-APPLICATION DATA:  
APPLICATION NUMBER: US 08/052,459







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compucon Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:07:10 ; Search time 1723.16 Seconds

(Without alignments)  
16197.975 Million cell updates/sec

Title: us-09-932-678-1

Perfect score: 2068

Sequence: 1

1 acacagagctgagctgagag.....tgacattggaatccacat 2068

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:  
1: em\_estha:  
2: em\_estim:  
3: em\_estlin:  
4: em\_estim:  
5: em\_estim:  
6: em\_estim:  
7: em\_estim:  
8: em\_estim:  
9: em\_estim:  
10: em\_estim:  
11: em\_estim:  
12: em\_estim:  
13: em\_estim:  
14: em\_estim:  
15: em\_estim:  
16: em\_estim:

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	815.4	39.4	1800	Y16973	RNY16973 Ka
2	675.4	32.7	788	H0502871	602550430
3	666	32.2	733	H1561523	603256222
4	651.6	31.6	996	BF203684	601866331
5	646	31.2	757	BF234354	601565596
6	637.4	30.8	1037	BF248305	602498983
7	617.8	29.9	671	AV703279	AV703279
8	614.6	29.7	736	H1464721	603202341
9	582.2	28.2	990	BF167061	602344740
10	551.4	26.8	828	H1761160	603043657
11	550.4	26.6	742	BF870692	601448528
12	548	26.5	550	AL597171	DKFZP314K
13	547.2	26.5	727	BF1549394	603140049
14	542.4	26.2	563	AW408066	U1-HF-BM0
15	536.8	26.0	564	AW588173	EST 370243
16	520.4	25.2	697	BF611364	602612725
17	517	25.0	535	AA213789	AA213789

18	513.8	24.4	835	10	BF611238	BF611238
19	488.4	23.6	498	10	BF110176	BF110176
20	487.4	23.6	497	10	BF221544	BF221544
21	482.8	23.3	498	10	BF149493	BF149493
22	476.6	23.0	503	9	AF179954	AF179954
23	471.8	22.8	628	9	BF642011	BF642011
24	470	22.7	490	10	BF502966	BF502966
25	465.8	22.5	496	10	BF535886	BF535886
26	463.2	22.4	632	10	BF614096	BF614096
27	461.4	22.3	1045	10	BF6297178	BF6297178
28	458	22.1	458	10	BF549643	BF549643
29	451.4	21.8	461	9	AF1928274	AF1928274
30	447.6	21.6	656	9	BF639379	BF639379
31	440.4	21.3	956	10	BF6295610	BF6295610
32	438.4	21.2	440	9	AA811628	AA811628
33	428.8	20.7	807	10	BF913245	BF913245
34	428.4	20.7	807	10	BF271211	BF271211
35	421.8	20.4	547	9	AA530643	AA530643
36	420.4	20.3	422	9	BF093077	BF093077
37	419.2	20.3	811	10	BF560866	BF560866
38	409	19.8	421	9	AA191111	AA191111
39	406.4	19.7	479	9	AA239267	AA239267
40	399.2	19.3	529	9	AA668124	AA668124
41	399.2	19.3	660	10	BF914307	BF914307
42	387	18.7	406	10	BF1018122	BF1018122
43	380	18.4	380	9	AA481295	AA481295
44	379.8	18.4	394	9	AA770928	AA770928
45	377.2	18.2	700	9	BF653390	BF653390

## ALIGNMENTS

RESULT 1  
Y16973  
LOCUS RNY16973 Rat Liver ESTs (E. Olivieri) Ratius norvegicus cDNA clone  
DEFINITION RNY16973 Rat Liver ESTs (E. Olivieri) Ratius norvegicus cDNA clone  
ACCESSION Y16973  
VERSION Y16973.2 GI:4727035  
KEYWORDS EST  
SOURCE Norway rat  
ORGANISM Ratius norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 1800)  
Olivieri, E., Soury, E., Ristler, L., Smith, F., Schneider, K., Lochner, K., Joudreau, J., Fey, G., and Salter, J. P.  
A novel set of hepatic mRNAs preferentially expressed during an acute inflammation in rat represents mostly intracellular proteins  
JOURNAL GENOMICS 57 (3), 352-364 (1999)  
MEDLINE 99263497  
COMMENT On Apr 2, 1998 this sequence version replaced q1:306069.  
CONTACT E. Olivieri  
DB INSERM  
543 chemin de la Breteque, 76213 Bois-Guillaume, France  
POLY(A-N).

## FEATURES

Source location/Qualifiers  
1..1800  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/date="1997-07-31"  
/clone="1997-07-31"  
/clone="1997-07-31"  
/issue="type=1110"  
/db\_xref="taxon:10116"  
/note="Organ: Liver; Vector: Lambda ZAP11; Library: construction: Ristler, J. A. et al. J. Biol. Chem. (1995) 270(50):29998-30006"  
BASE COUNT 473 a 393 c 433 g 501 t  
ORIGIN

1174

—

1234

107

3

16,34

1601A

—

1621

1680

MAK 2001











```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/cclone="IMAGE:526H011"
/clone_1lib="NH_MGC_97"
/lab_host="DH10B"
/notes="Organ: testis; Vector: pBluescriptR (modified
           pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (Gteqaa
           ); Oligo-dT primed using primer 5'-TTTCTTTTCTTTTCTT-3',
           size-selected for average insert size 2-2 kb and
           normalized to 10T 5. This is a primary library enriched
           for full-length clones and constructed using the
           Cap-Trapper method (Carlnici, in preparation). Library
           constructed by M. Brownstein (NIMH/NIHRI, National
           Institutes of Health). Note: This is a NH_MGC library."
```

[illegible]















of the  $2 \times 2$  matrix  $\mathbf{A}$  is

and the following table:

A vertical gel electrophoresis image showing a single dark band in the lane labeled '1'.

100

[illegible]

14 185013

4.4.4) ATR-FTIR spectra

184 Spillberg

6,400,000,000

[illegible]



[illegible]

seq name: /STDS/ncgdata/genecseq/genecseq.cml/AA2001A.DM1:AA01562

seq1\_documentation\_block:

ID: AA01562 standard: cDNA: 1421 bp.

AA01562:

18-JUL-2001 (first entry)

Human secretory molecule cDNA spm #52.

Human secretory molecule spm: spm library screening; gene therapy; cell signaling; cell proliferative disorder; attherosclerosis; connective tissue system disorder; AIDS; neurological disorder; Alzheimer's disease; nervous system disease; mental retardation; developmental disorder; neuromuscular disorder; microarray; Inocyte ID number: 42874526-c; ss

Human sapiens:

W0200125558 AZ.

05-APR-2001.

19-SEP-2000. 2000M-0525610.

26-SEP-1999. 990IS-0196624.

28-SEP-1999. 990IS-0196625.

02-OCT-1999. 990IS-0168611.

02-OCT-1999. 990IS-0168613.

02-OCT-1999. 990IS-0168614.

(IN.V) INYTE GENOMICS INC.

Hudson DM, Lincoln SE, Russo PD, Spito JA, Banville SC, Bratton SK, Dutoit CE, Cohen BJ, Rosen BD, Shah P, Chalup MS, Billam JL, Jones AL, Yu CY, Greenwell LB, Panzer SK, Roscherry AM, Wright RJ, Chen W, Liu TF, Yap PE, Stockhofner JK, Ashby S, Ford WT.

W01: 2001 256134/26.

New secretory polypeptides (spm) and the polypeptides they encode, for use in inducing antibodies and screening libraries of compounds.

Glenn L; Page 152; 161pp; English.

The present sequence for human secretory molecule cDNA spm #52 (Inocyte ID number 42874526-c) is 1 of 64 novel spm cDNA sequences (AA01511-AA01573) which encode for the secretory polypeptides spm. The spm polypeptides are useful for screening a compound for effectiveness in altering expression of a target polypeptide, where the target polypeptide comprises spm. Spm is also useful in a method for assessing the toxicity of a test compound. Spm and its fragments or complementary sequences, may be used to identify the presence of and/or determine the degree of similarity between two nucleic acid sequences.

Spm can also be used for a variety of diagnostic and therapeutic purposes, e.g., diagnosing a particular condition, disease or disorder associated with cell signaling, such conditions include cell proliferative disorders such as attherosclerosis, and cancers including leukemia, an immune system disorder e.g., acquired immunodeficiency syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's disease, nutritional and metabolic disease of the nervous system, mental retardation and other developmental disorders. Spm can also be used to design probes or antibodies for use in diagnostic assays, which may be used to monitor the progress of certain disorders associated with abnormal levels of expression of spm. In addition spm encoding spm may be used for somatic or gene therapy, for inducing antibodies, or in microarrays.

Sequence 1423 bp; 417 A; 247 C; 320 G; 449 T; 0 other;

of figure 1, seqs:

Quality: 1973.00 Length: 488  
 Ratio: 4.995 Gaps: 2  
 Percent Similarity: 80.943 Percent Identity: 80.533

## Alignment block:

US-09-932-678-2 x AAS01562

Align seq 1/1 to: AAS01562 from: 1 to: 1423

```

31 11ESerAsnMetArgAlaIleuGluIAsnAspPhePheAsnSerProProArg 47
|||||
194 AITTCGAAATATCCGICCATAGAGAAATGATTTTTCATCTCTCCCAAG 243
|||||
47 qySThrValArgPheGlyGlyThrValThrGluValIleuLeuLysTyrL 64
|||||
244 AAAACCTGTTCGGTTTGSTGAACTGTGACAGAGCTCTGCTGGAATACA 293
|||||
64 ySLysGlyGluThrAsnAspPheGluLeuLeuLysAsnGlnLeuLysAsp 80
|||||
294 AAAACGGTGAACAAATCACTTGAAGTTGTGAAGAACCAAGCTGTAGAT 343
|||||
81 ProAspIleLysAspAspGlnIleLeuAsnTrpLeuLeuGluPheArgSe 97
|||||
444 CGACGATTAAGATGACACATCATCACTGCTGCTAGCAATTCGTTG 493
|||||
97 rSerIleMetTyrLeuThrLysAspPheGluGlnLeuIleSerIleIle 114
|||||
394 TTCTATCATGTAAGTGAACAAAGCTTTGAGCAACTATCATATATAT 443
|||||
114 euArgIleProTrpIleAsnAspSerGlnThrValGluGlyIleu 130
|||||
444 TAAAGATGCTGTTGATATGAGTCAACACAGTACAGACAGATATTC 493
|||||
131 AlAPheLeuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProG 147
|||||
494 GCTTTCTGCTGAATCTGATCAACAGACAGCTTTTCTTCAGACCGTG 543
|||||
147 sLysSerMetIleAlaSerHisPheValProProArgValIleIleLys 164
|||||
544 TCCAGCATGATGCTCTGCTGCTTTTGCTCTCCCAAGTATCATTAAG 593
|||||
164 lucIAspValAspValSerAspSerAspAspGlnAspAspAsnLeuPro 180
|||||
594 AAGGCGATGTACATCTTTCAGATTCGATGATGAAGATGATATCTCT 643
|||||
181 AlaAsnPheAspThrCysHisArgAlaLeuGlnIleIleAlaArgTyr 197
|||||
644 GCAATTTTTCACACATGTCCACAGCCTTCCCAATATATAGCAGATAT 693
|||||
197 lProSerThrTrpTrpPheLeuMetProIleLeuValGluLysPheTrp 214
|||||
694 ACCATTCGACACGGTTCATTCATGCAATACATGCTGGAATAATTCAT 743
|||||
214 heValArgLysSerGluArgThrLeuGlnLysTyrValHisAsnLeuLeu 230
|||||
744 TTGTTGGAATAATCAGACAGACACTGCAATCTTACGTTCACTTAC 793
|||||
231 ArgLysSerValTyrPheProThrLeuArgHisGlnIleLeuGlnLeu 247
|||||
794 ACCATTCAGATATATTTTCCAACTGAGGCATCAAAATCTGGAGCTAT 843
|||||
844 TATTCGAAATCTACTCAGGTTCGATTCGATTCATCTCCGCGCAGAGAT 893
|||||
264 lAspAlaGlnGluThrAlaThrGlnThrCysGlyGlyThrAspSerThr 280
|||||
894 AAGATCTCTAAGAAACAGCAATCAAACTCTGCTGCAACAGATTCACG 943
|||||
281 GluGlyLeuPheAsnMetAspGlnAspGlnGlnThrGlnHisGlnThrL 297
|||||
944 GAAGGATGTTTATATG..... 961

```

```

297 SALaGlyProGlnArgLeuAspGlnMetValHisProValAlaIleuArg 414
961 ..... 961
314 euAspIleLeuMetSerLeuValIleuSerTyrMetLysAspValCysLys 340
961 ..... 961
431 ValAspGlyLysValAspAsnGlyLysTyrLysAspLeuTyrAlaAspLe 447
961 ..... 961
347 uLleAsnIlePheAspLysLeuLeuLeuProIleHisAlaSerCysHisV 364
961 ..... 961
364 aGlnPhePheMetPheTyrLeuCysSerPheLysLeuGlyPheArgAct 480
962 ..... 973
381 Ala-PheLeuGlnHisLeuArgLysLysLeuGlnAspProSerAspProA 397
|||||
924 GCGATTTTGSAAACATCTTTGSAAMAATTCGAGATTCATATATCTTG 1024
|||||
397 lAlIleArgGlnAlaAlaGlyAsnTyrTleGlySerPheLeuAlaArg 413
|||||
1024 CCATATATAGCAAGCTGCTGGAATATATATTCGAAACATTTTGTGA 1073
|||||
414 AlaLysPheIleProLeuIleThrValLysSerCysLys euAspLeuVal 440
|||||
1073 GCTAATTTATTTCTTATTATCTGTAAGAAGCTAAGCTACAGATCTT 1124
|||||
440 lAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrLysAla 447
|||||
1124 TAACTGATGAGCAATAATACCTTAATACCTGAGATTCGCAACAAAGCA 1174
|||||
447 heCysAspValAlaIleuHisGlyProPheTyrSerAlaCysGlnAlaVal 463
|||||
1174 TCTGATGATGTTGCTTCCATGAGCAATTTATTCAGCTGCGCAAGCTGG 1224
|||||
463 PheTyrThrPheValPheArgHisLysGlnLeuLeuLysSerGlyAsnLeu 480
|||||
1224 TTCATACACCTTTGTTTTCAGACAAACACACCTTTGAGCGCAACCTG 1274
|||||
480 sGlnLysLeuGlnTyrLeuGlnIleSerLeuAsnPheGluArgIleValMet 497
|||||
1274 ACAAAGTTTTCAGATATCTTCAGAGTTCGATTTTGAGCGGATAGATG 1324
|||||
497 eGlnLeuAsnProLeuLysLleCysLeuProSerValValAsnPhePhe 513
|||||
1324 GTCAGCTAAATACCTTGAAAGATTTGCTGCTGCTGCAATGATATTTT 1374
|||||
513 AlaAlaIleThr 517
|||||
1374 GTCGCAATCAACA 1385
|||||
seq_name: /stid1/seqdata/geneseq/geneseq-rmb1/MA2001A.DAT:AAS44584
seq_documentation_block:
ID AAS44584 standard: DNA: 1498 bp.
XX
AC AAS44584:
XX
DT 18-DEC-2001 (first entry)
XX
DE Human full-length polynucleotide sequence #9.
XX
KW Mammal: human; Rhesus monkey; baker's yeast; fission yeast; Norway rat;
mouse; Chinese hamster; African clawed frog; fruit fly; dog; Japanese
cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
nervous system disorder; inflammatory disorder; cell differentiation; ds;
anorexia; stem cell growth factor; activin; inhibin; cartilage; bone;
genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW

```









```

67  vllhrAsnAspPheGluLeuLeuLysAsnGlnLeuLeuAspProAspIleL 84
|||||
201 AACAAATGACTTGATGTTGTAAGAACGACGCTGTAATCAGCAACATAA 250
84 ys.....AspAspDlleLleAsnIrpLeuGlnPheArqSerSer 98
|||||
251 AGGTGGTGTGACCCGACATGCGATGATCTGAGGAGGAGGCGCAT 400
99 lle.....MetTyrLeuThrLysAs 105
|||||
301 GTGGCCGAGTGGCTTTCATCCCTTCGTGGCTTTCAGTTCGTCGCA 450
105 pHeGlnGlnLeuLleSerIleLeuArqLeuProIrpLeuAsnArqS 122
|||||
351 CCAAGTGGCTCAGCCGAAAGTTGCTAGCTGCTGCGGTGGGTCG 399
122 eGlnIrpValValGlnGlnIrpLeuAlaPheLeuGlnLysLeuValSer 138
399 ..... 399
139 AlaGlnIrpValPheLeuArqProGlySerMetIleAlaSerHisP 155
|||||
400 .....CCGCTGGCTGTGGGCGACATACAGCCCGACGCTCGCTAC 441
155 eValIrpProArqValIleIleLysGlnGlnLysValAspValSerAspS 172
441 ..... 441
172 eTAspAspGluAspAspGlnProAlaAspPheAspThrCysHisAr 188
|||||
442 .....AACCCGACAGCCGACGCGATTTGGCTGACTAT 478
188 gAlaLeuGlnIleLleAlaArqTyrValProSerThrProIrpPheLeu 205
|||||
479 .....CTGCTTGGCAGG.....ACACCGTGGTCTCGCA 506
205 eIrpIleLeuValGlnLysPheProPheValArqLysSerGlnValTyr 221
|||||
507 TCGCAATGACTGTGGAATAATTTCCATTGTCGMAATCAGAGAACAA 556
222 LeuGlnCysTyrValHisAsnLeuLeuArqLleSerValTyrPheProTh 238
|||||
557 CTGCAATGTACGTTCATTACTTAACCATTAAGTATTAATTTCCAC 606
238 rLeuArqHisGlnIleLeuGlnLeuIleIleGlnLysLeuLeuVal 255
|||||
607 CTTGAGGCATCAATTCGACCTTATTATTGAAAACTAAGTCTG 656
255 spValAsnAlaSerArqGlnGlyIleGlnAspAlaGlnIrpValArq 271
|||||
657 AATGAATGCAATCCCGACGGATATGACATGCTTAAGCAACACGCAAT 706
272 GlnThrCysGlnGlyThrAspSerThrGlnGlnLysPheAsnMetAsp 288
|||||
707 CAAACTGCTGCTGACGACATTCACGAGAAAGATTCGTAAATATGCT 756
288 uAspGlnGlnIrpGlnHisGlnIrpLysAlaGlnIrpGlnArqLeuAsp 305
757 ..... 757
305 lMetValHisProValAlaGlnArqGlnAspIleLeuMetSerLeuVal 321
|||||
758 .....ACATTATTAAAGCAAGTGGAG 778
322 LeuSerTyrMetLysAspValCysTyrValAspGlyLysValAspAsn 338
|||||
779 AICAAAGTTATC..... 790
338 yLysThrLysAspLeuTyrArqAspLeuIleAsnIlePheAspLysLeu 355
|||||
791 .....ATAATCAAAAGCTGCAAGACCTAC 815

```

```

455 eGlnProIrpHisAlaSerCysHisValIleIrpPheMetPheTyrLeu 471
|||||
816 TGCTG..... 820
472 CysSerPheLysLeuGlnIrpPheAlaGlnIrpPheLeuGlnIleL 488
|||||
841 .....CTCACTTTGATTCGACAGGCAATTTTGGAACTCTTGGAA 864
380 sLysLeuGlnAspProSerAspProAlaIleIleArqValAlaLac 404
|||||
865 AAAAGTCGAGGAAATGAACTAATCTGTCATCATCAACGACGCTG 914
405 AsnTyrIleGlnSerPheLeuAlaArqAlaLysPheIleProLeuIle 421
|||||
915 CATTAATATGAACTTTTGGCAAGACTACATTATTCTCTATTAC 964
421 rValLysSerCysLeuAspLeuLeuValAsnIrpLeuHisIleTyrLeu 448
|||||
965 TGTAAMACCATGCCATGATCTTTTGGTAACTGGCTGCAATATACCT 1014
448 sAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 454
|||||
1015 AIAAGTCAGATTCGCGAACAAGGCAATTTGATGATGTTGCTGCTCA 1064
454 lArProPheTyrSerAlaCysGlnAlaValIleTyrThrPheValPhe 470
|||||
1075 GACCATTTTACAGCTGTGGCAAGCTGTGTGTACACCTTGTCTTTA 1114
470 eLHisLysGlnLeuLeuSerGlnLysGlnGlnLysGlnGlnIleTyr 485
|||||
1115 GACACGACGACACTTTTGACGCGAAGCTGAAAGAGGAGGTGATG 1162
seqname: /SIN1/seqdata/genome/seq/genesetp-emb1/MA2001B.101:ANSR76.0
seq_dbox:meditation_block:
ID ANSR76.6 standard: cDNA: 1017 bp.
XX
AC AA:87636;
XX
XX
D7 1-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #23440.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
MC200175067-A2.
XX
XX
PD 11-MAR-2001.
XX
XX
PF 30 MAR-2001; 2001MO-US086.k1.
XX
XX
PR 31-MAR-2000; 2800US-0540217.
XX
PR 2-AUG-2000; 2000US-0649167.
XX
XX
PA (HSE-) HYSED INC.
XX
PI Brnasec RT, Liu G, Tang YL;
XX
XX
DR WP1: 2001-639362/73.
XX
DR P-PSDB: ARK23445.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1: SEQ ID No 23440; 107pp; English.
XX
CC The invention relates to isolated polynucleotide (1) and
CC polypeptide (1) sequences. (1) is useful as hybridisation probes.

```







560 PheProPhe.AsProCysVal.LeuLysArgSerLysLysPhe.IleAsp 575  
 401 TTCCCTTGATCCCTGCTGTGCGAAGAGGTCAAAAGATTCATCAT 450  
 576 ProlLe.Tyr.GluVal.Tyr.GluAspMetSerAlaGluGluLeuGln 591  
 451 CCTATTATACAGGATATGGGAGACATGAGTGTCTAAGAGCTACAGGA 500  
 591 uphElyLysProMetLysLysAspIleVal.GluAspGluAspAspAsp 608  
 501 GTTCAGAAACCCAGTGAAGAGACATAGTCAAGATGAAAGATGATGAC 550  
 608 IleuLysGlyGlyVal.ProGlnAsnAspThrVal.IleGlyIleThrPro 624  
 551 TTCTGAAGAGCAAGTCCCGCAGAAATACCGTATGGGATCAGCA 600  
 625 SerSerPheAspThrHisPheArgSerProSerSerValLysSerPr 641  
 601 AGCTGCTTGACACCCATTTCCGAGCTCTTCAAGATGATGCTGCTCCG 650  
 641 ofProValLeuTyr.MetGlnProSerProLeu 651  
 651 ACCCGCTTGTACATCGCAACGACGCTCCCTC 682

seq\_name: /STUS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AA592253

seq\_documentation\_block:  
 ID AA592253 standard: cDNA: 776 BP.  
 XX  
 AC AA592253;  
 XX  
 TT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #28057.  
 XX  
 KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 XX  
 PN W0200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001MO-US086-11.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INT.  
 PI Urmanac RT, Lin C, Tang YI;  
 XX  
 DR WP1: 2001-639462/73.  
 DX P-PSDB; AB028066.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 1: SEQ ID No 28057; 103pp; English.  
 CC  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC are to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AA564197-AA594764 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from Wipo  
 CC atftp.wipo.int/pub/publ/indirect\_sequences.

XX  
 XX Sequence 776 BP: 206 A; 157 C; 198 G; 215 T; 0 other:

aligner1\_scores:  
 quality: 882.00 length: 146  
 Ratio: 4.717 gaps: 5  
 Percent Similarity: 95.408 Percent Identity: 94.898

aligner2\_block:  
 US-09-932-678-2 x AA592253

Align seq 1/1 to: AA592253 from: 1 to: 776

1 MetAlaAlaProLeuLeuHisIleArg.LeuProGlyAspAlaAlaAs 17  
 196 ATGGCGGCGAGCGCTGCTTCAACATCTTTGTGGAGAAATGCGAGGCTT 247  
 17 SerSerAlaValLysLysLeuGlyAlaSerArgThrGlyIleAsnAs 44  
 248 CTCCTCTGTCAGTCAAGACGCTGTTCAGCTGGAGAGGAGATTTGAAA 297  
 32 MetValAlaLeuGluLysAsnAspIlePheAsnSerProProArgLysIleVal 50  
 298 TATGGTGATATACGAATGATTTTTCGAAATCTCTCCGCAACAAAAACCTG 447  
 50 AlaArgPheGlyGlyIleValThrGluValLeuLeuLysTyrLysLysGly 66  
 348 TTGGGTTTGGTGGAACTGTGACAGACGCTTGTCTGAAGTACAAAAGGCT 397  
 62 GluThrAsnAspPheGluLeuLeuLysAsnGluLeuLeuAspThrAsp 83  
 392 GAACAAATGCTTTGAGTTCCTTGAAACAGCAGCTGTGATATTCACAAAT 447  
 83 ElyAspAspGlnIleLeuAsnTyrLeuLeuGluPheArgSerSer 99  
 448 AAGGATGACGCAATATCAACTGGGCTGTCTAGAAATTCGCTTCTCT 497  
 95 MetTyrLeuThrLysAspPheGluGluLeuLeuLysSerIleLeuValAla 114  
 498 CATGTACTTGAAAGAGCTTTGAGCAACTTATACAGATTTAATGATAT 547  
 114 LeuProTyrPheAsnArgSerGlnThrValValGluGluTyrLeuAla 142  
 548 TSCCTTGGTTGAAATACAAAGTAAACAGTAAAGTAAAGTAAAGTAAAGT 597  
 142 LeuGluLysAsnGluValSerAlaGluIleValIleLeuValProLys 149  
 598 TCTTGGTAAATCTTATCAGCAGACAGCTTTCCTTACACGCTCTCTCA 647  
 149 GlnMetIleAlaSerHisPheValProProArgValIleIleLysGluGly 165  
 648 GATATGATCTTCTCCATTTTGTGCTTCCGCACTGATGATTAAGCAAGAG 697  
 165 AspValAspValSerAspSerAspAspGluAspAspAspMetProValAs 182  
 698 ATGTAGATGTTTCGATTTGATGATGAAAGATGATATATCTTCTGCAAA 747  
 182 PheAspThrLysHisArgAlaAlaGlu 191  
 748 TTTGACAAATCTCAGACAGACGCTTGTCAA 775









79	caaspspslpslllelysspsps-glulelileasntpleuengluph	95
103	ltaatccagaccatlaagggatlaaccacatcatatnactggatgtatnaatt	152
95	evasrserllemoltyrleuthr-lysasphegln-glulellesl	111
153	cccttcttcgcacatgactgcacacaaacacattgagccacactatcag	202
111	rllelleleuagluaprotprlucnaasatgsergln-thrvalylolu	127
203	tattatattgca-nattggcttggttgatatacagctcaaacagatagtgaa	252
128	glnttytleualaphelengl-ysasleuvalserlaglnthrvalphele	144
253	gacatttggtctttcttggaatccttgatcagcattaaacggttttct	302
144	uagprcocytleu-sermetlilaaser-hlsphelapropoargval	160
303	cagacgcctgctgcacatgattgcttcccatatttgccctccgcagng	352
161	llellelysglgl-vaspvalaspsvalseraspsaspsaspalnasps	177
353	atcatttaagaaagggcagatcagatgatttccatatttgatgatgaaatcca	402
177	pasluclproalaasnphesaspthr-cysllsargalaleuglnlillea	194
403	tattcttccgtgaaattttgacacatatacmaacagccttgcaaatatag	452
194	laartrtyvalproserthrprotrpphe	203
453	caaaatagaccatcatatgactatatttt	481



0M of: 05-09-932-678-2 to: N:geneseq\_012802.\* out\_format: p1s

Date: Jul 30, 2002 7:44 AM

About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-Model: frame\_gpn\_model -DEV: x1h  
-Query: 05-09-932-678-2 -USPRO: /usr932678/runatc\_29072002\_155156\_19107/app\_query.fasta\_1.717  
-DB: N:geneseq\_032802 -QPM: fastlap -SUFFIX: 011.fng -GAPOP: 4.500  
-GAPEXT: 0.050 -MINMATCH: 0.100 -LOCPL: 0.000 -LOCPEXT: 0.000  
-GAPOP: 4.500 -GAPEXT: 0.050 -XGAPOP: 60.000 -XGAPEXT: 60.000  
-XGAPOP: 6.000 -XGAPEXT: 7.000 -XGAPOP: 60.000 -XGAPEXT: 60.000  
-DELOP: 6.000 -DELEXT: 7.000 -START: 1 -MAPR1X: 01190  
-TRANS: human40.cdi -LIST: 45 -DOCALLIGN: 200 -THR: score-quality  
-HRR\_MIN: 1 -ALLEN: 15 -MODE: LOCAL -OUTPR: p1s -MOR: ext  
-HEPSTLZ: 500 -MINLEN: 0 -MAXLEN: 2000000000  
-USPR: 0509912678 -CGEN: 1.0 -NCP: 6 -ICP: 3 -LONGLOG  
-DEV: TIMEOUT=120 -WARN: TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: 05-09-932-678-2

Query length: 651

Database: N:geneseq\_032802.\*

Database sequences: 1736436

Database length: 859457221

Search time (sec): 212.800000

WARN: XGAPOP and XGAPEXT must be equal. Assuming XGAPOP=60.000  
WARN: XGAPEXT and XGAPEXT must be equal. Assuming XGAPEXT=60.000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AAA98384 +		463.00	8547.56	0.0	20
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AAA98384 +		295.00	5435.62	2.0e-294	1
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AACT6937 +		276.00	5083.83	8.0e-275	14
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AACT6937 +		240.00	4415.46	1.2e-247	1
/SID1/gcgcdata/geneseq/geneseq-emb1/NA1999.DAT:AAV90248 +		97.00	1776.37	1.3e-90	34
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS87641 +		97.00	1759.62	1.2e-89	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS73660 +		93.00	1685.90	1.5e-85	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS91952 +		93.00	1685.90	1.5e-85	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS91952 +		79.00	1436.52	1.1e-71	7
/SID1/gcgcdata/geneseq/geneseq-emb1/NA1999.DAT:AAAS1663 +		75.00	1366.09	8.9e-68	43
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS92248 +		75.00	1364.49	1.2e-67	5
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS92248 +		71.00	1277.64	8.1e-63	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS81038 +		70.00	1264.93	4.1e-62	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS87636 +		69.00	1249.12	3.1e-61	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS92255 +		66.00	1187.02	9.0e-58	2
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS80743 +		63.00	1144.98	2.0e-55	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS91986 +		63.00	1144.98	2.0e-55	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS91986 +		63.00	1144.98	2.0e-55	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS92476 +		62.00	1097.85	8.4e-53	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS92476 +		62.00	1092.92	1.6e-52	3
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK85168 +		61.00	1095.36	1.2e-52	2
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AACT7798 +		59.00	1067.05	4.4e-51	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:AAAS80743 +		59.00	1062.13	8.2e-51	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK68674 +		58.00	1028.61	6.0e-49	2
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK68674 +		58.00	1021.51	1.5e-48	2
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK85168 +		54.00	974.61	6.1e-46	6
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK85168 +		45.00	809.59	9.5e-37	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AACT7798 +		43.00	775.18	7.9e-35	34
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AACT1901 +		48.00	686.21	7.1e-30	21
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK44215 +		35.00	625.60	1.7e-26	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK44215 +		35.00	625.60	1.7e-26	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK44215 +		35.00	625.60	1.7e-26	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK44215 +		35.00	625.60	1.7e-26	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK30411 +		35.00	625.60	1.7e-26	4
/SID1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK30411 +		35.00	625.60	1.7e-26	4

seq\_name: /SID1/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:AAA98384  
seq\_documentation\_block:  
ID: AAA98384 standard: cDNA: 2040 bp.  
AC: AAA98384:  
XX  
XX 08-PEB-2001 (first entry)  
XX  
XX Human RNA polymerase I transcription factor TIF-1A cDNA  
XX  
XX RNA polymerase I transcription factor TIF-1A; and tumor treatment;  
XX antiproliferative cell proliferation; cancer; tissue regeneration; ss.  
XX  
XX Homo sapiens.  
XX  
XX W020025316-A1.  
XX  
XX 21-SEP-2000.  
XX  
XX 08-MAR-2000; 2000W0-DE90767.  
XX  
XX 17-MAR-1999; 90GE-1011942.  
XX  
XX (DEKR-) DEUT KREHSFMSCHONSZENTKUM  
XX  
XX Chromat 1, Vindition M;  
XX  
XX WPI: 2000-587527/55.  
XX P FISH: AAB10946.  
XX  
XX Nc. DNA encoding the transcription factor TIF-1A, useful for prevention  
XX of treating diseases associated with abnormal cell proliferation,  
XX particularly tumors.  
XX  
XX claim 1; Fig 2; 38pp; German.  
XX  
XX This invention describes a novel DNA sequence (1) that encodes the RNA  
XX polymerase I transcription factor TIF-1A which has antitumor;  
XX antiproliferative and proliferation-inducing activity. The invention also  
XX describes (1) DNA (1a) encoding a protein (11) with the biological  
XX activity of TIF-1A; (2) a ribozyme (R) corresponding to (1) or (1a) and  
XX able to bind specifically to, and cleave, its transcribed RNA so as to  
XX reduce or inhibit synthesis of the corresponding protein; (3) an  
XX antisense RNA (AS) with binding properties similar to R; (4) an  
XX expression vector that contains (1), (1a) or sequences that encode R or  
XX AS; (5) host cells containing the vectors of (4); (6) TIF-1A or (11)  
XX encoded by (1) or (1a); (7) preparation of TIF-1A or (11) by culturing  
XX cells of (6); (8) ligands that bind to TIF-1A or (11); (9) antagonists  
XX that weaken or block the activity of TIF-1A or (11); (10) a diagnostic  
XX method for detecting abnormal TIF-1A expression; and (11) kit for  
XX carrying out the method in (10). (1), and similar sequences that encode  
XX proteins with equivalent activity, expression vectors containing them, or  
XX their expression products are used to treat or prevent disorders  
XX associated with reduced cellular proliferation, to stimulate cellular  
XX proliferation, and to promote tissue regeneration, e.g. after injury or  
XX radiation therapy. Ribozymes, antisense sequences directed against (1),  
XX also ligands associated with excessive cellular proliferation and to inhibit  
XX proliferation, especially in treatment of cancers, (1) and specific  
XX ligands for TIF-1A (particularly antibodies (Ab)) are also useful for  
XX diagnosis of altered TIF-1A expression by (indirect determination of the  
XX concentration, length and/or sequence of TIF-1A or its mRNA, or a for  
XX detecting mutations. Ab can also be used for immunoprecipitation of  
XX TIF-1A and for isolation of related sequences from cDNA expression









```

341 SASPLFNTYArqAspIcoulEasulIephEasplysLeuLeuLeuProt 438
      |||||||
423 GGATCATATATGGAGCTGATAAACATCTTGATAAACCTCCTTGCCCA 472
      |||||||
358 hrhhsaiaserCysHsIvalGlnPhePheMetPheTyrLeuCysSerPhe 374
      |||||||
473 CCCTATGCC TCTGTCATGACAGTTTTCATGTTTACTCCTGCTAGTTTC 522
      |||||||
375 LysLeuGlyPheAlaGluAlaPheLeuGlnHisLeuTyrPlyLysLeuGln 391
      |||||||
523 AAATGGGATTCGATAGAGCCATTTTCGACACATCTCTGAAAAAATGCA 572
      |||||||
391 nAspPProSerAspProAlaIleIleArgGlnAlaAlaCysAsnTyrIleG 408
      |||||||
573 GGATCCAGCAATATCTGTCATCATCAGCAGGCTGCTGAAATATATATG 622
      |||||||
408 LysSerPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSer 424
      |||||||
623 GAACCTTTTCGCAACAGCCAAATATATCTCTTATTAACCTAAATCA 672
      |||||||
425 CysLeuAspLeuLeuValAsnTyrPheHisIleTyrLeuAsnAsnGlnAs 441
      |||||||
674 TGGTAAACATCTTTGGTAACTGGTGGCACAATATACCTTAATACCAACA 722
      |||||||
441 pSerGlyThrLysAlaPheCysAspValAlaLeuHisClyProPheTyrS 458
      |||||||
723 TTTCAGCAACAAGCCATTTTCGATGTTGCTCTCATGAGCCATTACT 772
      |||||||
458 cTAlaCysGlnAlaValPheTyrThrPheValPheArgHisLysGlnLeu 474
      |||||||
774 CAGCTTCGCAAGCTCTCTCTACAGCTTGTTGTTTACGACACACAGCTT 822
      |||||||
475 LeuSerGlyAsnLeuLysClyLeuGlnTyrLeuGlnSerLeuAsnPh 491
      |||||||
823 TTTCATCGAAACCTGAAAGAGGTTCGATATCTCTTCACTCGAATTT 872
      |||||||
491 ecIArgIleValMetSerGlnLeuAsnProLeuLysIleCysLeuProS 508
      |||||||
873 TGACGCGATAGTGTATGACGACGTAATCCCTCAACATTTCCCTGCTCT 922
      |||||||
508 cTValValAsnPheAlaAlaIleThrAsnLysTyrGlnLeuValPhe 524
      |||||||
923 CAGTGGTAACTTTTCTGCAATCCAAATAACTACCACTGCTGCTTC 972
      |||||||
525 CysTyrThrIleIleGlnArgAsnAsnArgGlnMetLeuProValIleAr 541
      |||||||
974 TGTACACCATCATTCAGAGCAACATCCGCAATGCTGCTGATCATTAG 1022
      |||||||
541 gSerThrAlaGlyClyAspSerValGln 550
      |||||||
1023 GAGTACCGCTGAGAGCAATCTCACTGACAG 1050
      |||||||
seq_name: /SUBS1/qcdata/geneseq/geneseq-emb1/NA2001A.DAT.AAS01562
seq_documentation_block:
ID AAS01562 standard: cDNA, 1423 BP.
XX
XX AAS01562:
XX
XX 18-JUL-2001 (first entry)
XX
XX DE Human secretory molecule cDNA spkm #52.
XX
XX KW Human: secretory molecule; spkm: SPTM: library screening; gene therapy:
XX cell signaling; cell proliferative disorder; atherosclerosis; cancer;
XX immune system disorder; AIDS; neurological disorder; Alzheimer's disease;
XX nervous system disease; mental retardation; developmental disorder;
XX neuromuscular disorder; microarray; incyte ID number 4287452dec; ss.
XX
XX OS Homo sapiens.
XX
XX PN W0200123558-A2.

```

```

XX
XX 05-APR-2001.
XX
XX 19-SEP-2000: 2000W0-0525610.
XX
XX 28-SEP-1999: 990S-0156624.
XX 28-SEP-1999: 990S-0156625.
XX 02-DEC-1999: 990S-0168611.
XX 02-DEC-1999: 990S-0168613.
XX 02-DEC-1999: 990S-0168614.
XX
XX (INCYTE) INCYTE GENOMICS INC.
XX
XX Hodgson DM, Lincoln SE, Russo ED, Spito PA, Rantville SC,
XX Bratcher SR, Munroe GE, Cohen RJ, Rosen RH, Shah P, Chai-P MS,
XX Hillman JL, Jones AL, Yu JY, Greenwalt LB, Panzer SG,
XX Roseberry AM, Wright RJ, Chen W, Liu IF, Yap PG, Stockbrecher JK,
XX Amesey S, Fong WJ.
XX
XX WPI: 2001-258134/25.
XX
XX New secretory polypeptides (sptm) and the polypeptides they encode,
XX for use in inducing antibodies and screening libraries of compounds.
XX
XX Claim 1: Page 152; 161pp: English.
XX
XX The present sequence for human secretory molecule cDNA spkm #52
XX (incyte ID number 4287452dec) is 1 of 63 novel sptm cDNA sequences
XX (AAS01511-AAS01573) which encode for the secretory polypeptides sptm. The
XX sptm polypeptides are useful for screening a compound for
XX effectiveness in altering expression of a target polypeptide, where
XX the target polypeptide comprises sptm. Sptm is also useful to a method
XX for assessing the toxicity of a test compound. Sptm and its fragments or
XX complementary sequences, may be used to identify the presence of and/or
XX determine the degree of similarity between two nucleic acid sequences.
XX Sptm can also be used for a variety of diagnostic and therapeutic
XX purposes, e.g., diagnosing a particular condition, disease or disorder
XX associated with cell signaling, such conditions include cell
XX proliferative disorders such as atherosclerosis, and cancers including
XX leukemia, an immune system disorder e.g., acquired immunodeficiency
XX syndrome (AIDS), a neurological disorder such as epilepsy or Alzheimer's
XX disease, nutritional and metabolic disease of the nervous system, mental
XX retardation and other developmental disorders, and muscular dystrophy and
XX other neuromuscular disorders. Sptm can also be used to design probes
XX useful in diagnostic assays, which may be used to monitor the progress of
XX conditions or disorders associated with abnormal levels of expression of
XX sptm. In addition sptm encoding SPTM may be used for somatic or germ-line
XX gene therapy, for inducing antibodies, or in microarrays.
XX
XX
XX Sequence 1423 BP: 417 A; 247 C; 320 G; 439 T; 0 others.
XX
XX
XX Alignment scores:
XX      Quality: 240.00      Length: 240
XX      Ratio: 1.000      Gaps: 0
XX Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment_block:
XX 08-09-92-678-2 x AAS01562
XX
XX Align seq 1/1 to: AAS01562 from: 1 to: 1423
XX
XX 31 |||SerAsnMetIArgAlaLeuGlnAsnAspPheAsnSerProLysAt 47
XX |||||||
XX 154 ATTCAAAATATGAGCATATAAAGATTTTTCATTTTCGCCCAAG 244
XX |||||||
XX 47 gYsThrValArgPheClyGlyThrValThrGlnValLeuLeuLysTyrL 44
XX |||||||
XX 244 AAAAAGTGTGGTGGTGGTGGCAACCTGACAGAACGCTTCTGAGAGTCA 294
XX |||||||
XX 64 ySLysGlyGlyThrAsnAspPheGlnLeuLeuLysAsnGlnLeuLeuAsp 80
XX |||||||
XX 294 AAAAGGTCGAAACAAATGATTTTGAGTTCGTAACAAACCTGATATAT 344

```



XX DE DNA encoding novel human diagnostic protein #23445.  
 XX XX  
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX OS Homo sapiens.  
 XX W0200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001: 2001WO-0808631.  
 XX 31-MAR-2000: 2000US-0540217.  
 XX 23-AUG-2000: 2000US-0649167.  
 XX (HSE-) HYSEQ INC.  
 XX Demanac RT, Liu C, Tanq YT;  
 XX WI: 2001-639362/73.  
 XX P-PSUB: AB023454.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX Claim 1: SEQ ID No 23445; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences; (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging or sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/publist/seq\\_sequences](http://wipo.int/pub/publist/seq_sequences).  
 XX  
 XX Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 other;  
 XX  
 XX Alignment scores:  
 XX Quality: 97.00 Length: 97  
 XX Ratio: 1.000 Gaps: 0  
 XX Percent Similarity: 100.000 Percent Identity: 100.000  
 XX  
 XX alignment\_block:  
 XX US-09-932-678-2 x AAS87641  
 XX  
 XX Align seq 1/1 to: AAS87641 from: 1 to: 3169  
 XX  
 XX 366 LeuTTPyLysLysLeuGluAspProSerAspProAlaIleIleArgAlaI 402  
 XX TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
 XX 857 GTTTCGAAAAAATTGCAGATCCAACTTATCTTCATATCAGACACACG 906  
 XX 402 AAlaGlyAsnTyrTleGlySerPheLeuAlaArgAlaLysPheLeuPol 419  
 XX TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
 XX 907 TCTTCGAAATTAATTCGAAGCTTTTTCGCAAGACCTAAATTATTCCTC 956

419 eulleTluValLysSerCysLeuAspLeuLeuValAsuTrrtrPolHisIle 435  
 436 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
 957 TTATTACTGAAATATCATCTGAGATGTTTTCCTTAACTGCTGCAACA 1006  
 436 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
 436 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
 1007 TACCTTAATTAACAGAGATTGAGGAAATAGAGATTCGAGATTCAT 1056  
 452 DHSGLYTPHPLCYTSRA)ACYSGLNAlaValPheTyrIlePheValP 469  
 469 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT  
 1057 CCATGACCACTTACTCAGCCGCAACCTGCTCTACACCTTCTCTT 1106  
 469 heatqHISLysGluLeuSerGlyAsnLeuLysGluGly 482  
 1107 TTACACACACATCAGCTTTTTCAGCGCAAGACTCAAAAGACGC 1147  
 seq\_name: /SIDSI/gcdata/gene/seq/gene/seq-emb1/MA2001B.LAT:MA574660  
 seq\_documentation\_block:  
 ID AAS74660 standard: cDNA; 2991 BP.  
 XX  
 AC AAS74660;  
 XX  
 XX 1: FEB-2002 (first entry)  
 XX  
 XX DNA encoding novel human diagnostic protein #9464.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX OS Homo sapiens.  
 XX W0200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001: 2001WO-0808631.  
 XX 31-MAR-2000: 2000US-0540217.  
 XX 23-AUG-2000: 2000US-0649167.  
 XX (HSE-) HYSEQ INC.  
 XX Demanac RT, Liu C, Tanq YT;  
 XX WI: 2001-639362/73.  
 XX P-PSUB: AB009473.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX Claim 1: SEQ ID No 9464; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences; (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging or sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human



XX 13-FEB-2002 (first entry)  
 DT  
 XX DNA encoding novel human diagnostic protein #28057.  
 DE  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 PN W0200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 30-MAR-2001: 2001WO-0508631.  
 XX  
 XX 31-MAR-2000: 2000US-0540217.  
 XX 23-AUG-2000: 2000US-0645167.  
 XX  
 PA (HYTE-) HYTEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DK WPI: 2001-619362/73  
 XX P-PSDB: ABG28066.  
 XX  
 PT New isolated polypeptide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 1: SEQ ID NO 28057: 103pp: English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences).  
 XX  
 SO Sequence 776 BP; 206 A; 157 C; 198 G; 215 T; 0 other;

Alignment\_scores:  
 Quality: 79.00 Length: 94  
 Ratio: 0.859 Gaps: 1  
 Percent Similarity: 98.925 Percent Identity: 98.925

Alignment\_block:  
 US-09-932-678-2 x AAS92253 ..

Align seq 1/1 to: AAS92253 from: 1 to: 776

100 MetTfLeuThrLysASPheGluGluLeuLeuSerLeuLeuLeuArgLe 116  
 |||||||  
 499 AGTGTCTGCACAAAGACTTTGAGCAATTCATGATATATATGAAAT 548  
 |||||||  
 116 uPro.TfPLeuAsnArgSerGlnThrValValGluGluTfTfLeuAlaPhe 132

|||||  
 541 GGCCTTGCTTGAAATGAGAGCTCAACAGTACGAGTACAGCACTATTTGGCTTTT 598  
 |||||||  
 133 LeuTfVAsnLeuValSerAlaGlnThrValPheLeuAlaGfProCysLeuSe 149  
 |||||||  
 559 CTTCGATATCTTGATCATACACAGATGCTTTCTGCAACGCTGTCTGCAAT 648  
 |||||||  
 149 MetTfLeuAsnSerHisPheValProTrpArgValIleIleLysGlnLys 166  
 |||||||  
 649 CAGATATGCTTCCCATTTTGTGCTCTGCTGCACTGACATTAAGCAAGTGT 698  
 |||||||  
 166 SPVALSPVALSPASPSPGASASPAGLSPASPASPSPSPSPSPSPSPSPSP 182  
 |||||||  
 659 AGTACAGATGTTTCAGATTCGATGATGAGATGATGATGATGATGATGAT 748  
 |||||||  
 183 PheAspThrCysHisAsnArgAlaLeuGln 191  
 |||||||  
 749 TTGCACACATGCTGCACAGCCTTGCA 775  
 |||||||

seq\_name: /STDS1/seqdata/geneseq/emb1/NA11999.DAT: AAS51664

seq\_documentation\_block:  
 ID AAS51664 standard: cDNA: 437 BP.  
 XX  
 AC AAS51663;  
 XX  
 DT 21-JUN-1999 (first entry)  
 XX  
 DE Human secreted protein 5' EST SEQ ID NO:242.  
 XX  
 KW Human secreted protein; EST; expressed sequence tag; diagnostic;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW TGF-beta; hormone regulation; chemokine; chemokine receptor; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W05905549-A2.  
 XX  
 PD 11-FEB-1999.  
 XX  
 PF 31-JUL-1998: 98WO-1801231.  
 XX  
 PR 01-AUG-1997: 97US-0905279.  
 XX  
 PA (c-EST) GENSER1  
 XX  
 PI DeCort A, Dumas M, Line Edwards J, Lacroix R;  
 XX  
 DR WPI: 1999-153779/13.  
 XX P-ESTDB: AAV12885.  
 XX  
 PT May nucleic acids encoding human secreted proteins obtained from  
 PT cDNA libraries derived from testis, ovary, uterus and spleen tissue  
 XX  
 PS Claim 1: Page 344-345: 522pp: English.  
 XX  
 CC AAS51459 to AAS51691 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAV12881 to  
 CC AAV12913, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell proliferation  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemokine/chemokine activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter



XX XX  
 PN W0200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 40-MAR-2001: 2001WO-0508631.  
 XX  
 PK 31-MAR-2000: 2000US-0540217.  
 XX 23-AUG-2000: 2000US-0649167.  
 XX  
 PA (HYPED) HYSEQ INC.  
 XX  
 PI Lumanan RL, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-649462/73.  
 XX P-PSNR: A8G28667.  
 DR  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, for genetic mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 1: SEQ ID No 28058; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from Wipo  
 CC at [http://wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).  
 XX  
 XX Sequence 3175 BP: 811 A; 849 C; 779 G; 736 T; 0 other:  
 SO  
 alignment\_scores:  
 Quality: 71.00 Length: 71  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-932-678-2 x AAS92254 ..  
 Align seq 1/1 to: AAS92254 from: 1 to: 3175  
 200 ThrProThrPhaIaMetProIleValGluValGluIuIuSphProPhaValA 216  
 491 ACACCGTGGTTCACATGCAATACCTGCTGCAAAATTCATCTTCTTCG 540  
 216 glysserGLuArGThrIleuGluGlyTyTAlaHisAsnLeuLeuArgIles 234  
 541 AAATTCAGAGAGAACACTGGAATGTTACCTTCATTAAGCATTA 590  
 233 eValTyTPhProThrLeuArGHisGluIleleuGluIleuIleleu 249  
 591 GGTATATTTTCACACTTGAGCATGAATTCGAGCATTAATATTATGAA 640  
 250 LysIeuLeuLysLeuAspValaAsnAlaSerATgGlnClyIleGluAspAl 266  
 [|||||

641 AACIACCTCAAGTCGATGCTGAATGATACCGGAGGCTATTGAAACATCT 690  
 266 aGluGluThrAla 270  
 691 TGAGAAACACCA 793  
 seq\_name: /S1DSI/seqdata/geneseq/geneseq-emb1/NA2001A.DAT:ABA83038  
 seq\_documentation\_block:  
 ID ABA83038 standard; DNA; 1461 BP.  
 XX  
 AC ABA83038;  
 XX  
 XX 05-FEB-2002 (first entry)  
 XX  
 DE Human transcription factor TRFX-65 coding sequence.  
 XX  
 KW Human; transcription factor; TRFX; cell proliferative disease;  
 KW autoimmune disease; inflammation; neurological disease;  
 KW developmental disorder; cancer; AIDS; infection; cytostatic; and HIV;  
 KW neuroprotective; antiinflammatory; gene therapy; ds.  
 XX  
 XX Hemo sapiens.  
 XX  
 XX W0200172777-A2.  
 XX  
 XX 04-OCT-2001.  
 XX  
 XX 13-MAR-2001: 2001WO-US08117.  
 XX  
 XX 13-MAR-2000: 2000US-0188986.  
 XX  
 XX (JMY-) INCYTE GENOMICS INC.  
 XX  
 XX Hillman JV, Baughn MR, Yue H, Lal P, Lu DM, Patterson C,  
 XX Azimzai Y, Handman O, Tang YT, Mathur V, Shah P, Au-Young J,  
 XX Reddy R.  
 XX  
 DR WPI: 2001-570896/64.  
 DR P-FED3: ABB50214.  
 DR  
 PT Novel transcription factor polypeptides, used to treat diseases  
 PT associated with altered activity and expression of TRFX, and to screen  
 PT for agents capable of modulating its activity.  
 XX  
 PS Claim 11: Page 299; 327pp; English.  
 XX  
 CC The present sequence is the coding sequence for a human transcription  
 CC factor. The transcription factor and its coding sequence are useful in  
 CC the diagnosis, treatment and prevention of diseases associated with  
 CC altered expression of the transcription factor e.g. cell proliferative,  
 CC autoimmune/inflammatory, neurological and developmental disorders. A  
 CC number of specific disorders/diseases are given in the specification,  
 CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
 CC allergies, anemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
 CC dermatitis, diabetes mellitus, emphysema, Crohn's disease, goitre,  
 CC Grave's disease, multiple sclerosis, osteoarthritis, psoriasis,  
 CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
 CC colitis, leishmaniasis, Alzheimer's disease, Huntington's disease, Parkinson's  
 CC disease, stroke, and viral, bacterial, fungal and protozoal infections.  
 XX  
 SO Sequence 1461 BP: 413 A; 335 C; 298 G; 415 T; 0 other:  
 alignment\_scores:  
 Quality: 73.00 Length: 70  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-932-678-2 x ABA83038 ..  
 Align seq 1/1 to: ABA83038 from: 1 to: 1461





XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 1: SEQ ID No 28059; 103bp; English.

XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from Wipo  
 CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

SO Sequence 2410 BP: 672 A: 437 C: 547 G: 754 T: 0 other:

alignment\_scores:  
 Quality: 56.00 Length: 66  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-932-678-2 x AAS92255 ..

Align seq 1/1 to: AAS92255 from: 1 to: 2410

427 AspleuLeuValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGI 443  
 |||||||  
 1 GATTTTGGTTACTGGCTGCCACATATACCTTATATACGAGATTCGGG 50  
 443 YThrLysAlaIlePheCysAspValAlaIleuHisClyProPheTyrSerAlac 460  
 |||||||  
 51 AACAAAGGCAATCTCGCATGTGCTCTCCATGGACCATTTTACTCAAGCT 100  
 460 YScIAlaValAlaPheTyrThrPheValPheArgHisLysGlnLeuLeuSer 476  
 |||||||  
 101 GCCAAGCTGTGTCTACACCTTTGTTTAGACACACAGCCTTTTCAGC 150  
 477 GlyAsnLeuLysGlnClyLeuGlnTyrLeuGlnSerLeuAsnPheGlu 492  
 |||||||  
 151 CCAAGCTGAAGAGCTTTGCAGTATCTTCAGAGCTGGAATTTGAG 198

•  
•  
•  
•

seq_name: /cgn2_6/prodata/1/ind/6h_COMB.seq:US-08-067-782A-3 *	7.00	1.0E+15	442.40	4.00
seq_documentation_block:				
Sequence 202, Application US/08905223				
Patent No. 622023				
GENERAL INFORMATION:				
APPLICANT: Edwards, Jean-Baptiste D.				
APPLICANT: Ducloux, Armand				
APPLICANT: La-toix, Bruno				
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS				
NUMBER OF SEQUENCES: 503				
CORRESPONDENCE ADDRESS:				
ADDRESSEE: Knobbe, Martens, Olson & Bear				
STREET: 501 West Broadway				
CITY: San Diego				
STATE: California				
COUNTRY: USA				
ZIP: 92101-3505				
SEQUENCER READABLE FORM:				
MEDIUM TYPE: PLASMID				
COMPUTER: IBM PC compatible				
OPERATING SYSTEM: Windows				
SOFTWARE: word				
CURRENT APPLICATION DATA:				
APPLICATION NUMBER: US/08/905,223				
FILING DATE:				
CLASSIFICATION: 536				
ATTORNEY/AGENT INFORMATION:				
NAME: Israelson, Ned A.				
REGISTRATION NUMBER: 29,655				
REFERENCE/DOCKET NUMBER:				
TELECOMMUNICATION INFORMATION:				
TELEPHONE: (619) 235-8550				
TELEFAX: (619) 235-0176				
INFORMATION FOR SEQ ID NO: 202:				
SEQUENCE CHARACTERISTICS:				
LENGTH: 409 base pairs				
TYPE: NUCLEIC ACID				
STRANDEDNESS: DOUBLE				
TOPOLOGY: LINEAR				
MOLECULE TYPE: cDNA				
ORGANISM: Homo Sapiens				
TISSUE TYPE: Brain				
FEATURES:				
NAME/KEY: other				
LOCATION: 53..84				
IDENTIFICATION METHOD: blastn				
OTHER INFORMATION: identity 90				
OTHER INFORMATION: region 82..112				
OTHER INFORMATION: id HS046355				
OTHER INFORMATION: est				
FEATURE:				
NAME/KEY: other				
LOCATION: 53..84				
IDENTIFICATION METHOD: blastn				
OTHER INFORMATION: identity 90				
OTHER INFORMATION: region 82..112				
OTHER INFORMATION: id HS046355				
OTHER INFORMATION: est				
FEATURE:				
NAME/KEY: other				
LOCATION: 227..276				
IDENTIFICATION METHOD: blastn				
OTHER INFORMATION: identity 94				



```

seq_documentation_block:
1 Sequence 9, Application US/09500569
2 Patent No. 6329204
3 GENERAL INFORMATION:
4 APPLICANT: Canon, Rebecca F.
5 APPLICANT: Katdalski, Antoni
6 APPLICANT: Shen, Jennie
7 TITLE OF INVENTION: Plant Galleic acid 3-O Methyltransferase Homologs
8 FILE REFERENCE: BB1327 US NA
9 CURRENT APPLICATION NUMBER: US/09/500.569
10 EARLIER FILING DATE: 2000-02-09
11 EARLIER APPLICATION NUMBER: 60/119,587
12 NUMBER OF SEQ ID NOS: 28
13 SOFTWARE: Microsoft Office 97
14 SEQ ID NO 9
15 LENGTH: 1342
16 TYPE: DNA
17 ORGANISM: Oryza sativa
18 US-09-500-569-9

alignment_scores:
1 Quality: 8.00 Length: 8
2 Ratio: 1.000 Caps: 0
3 Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-932-678-2 x US-09-500-569-9/rev ..

Align seq 1/1 to reverse of: US-09-500-569-9 from: 1 to: 1342

      14 AAAAAAAAAAASerSeraIval 21
         |||
        235 GGCGCCGCCACCTGTCGGCAGC 212

seq_name: /cgen2_6/plodata/1/ina/bA_COMB.seq:US-08-815-809-2

seq_documentation_block:
1 Sequence 2, Application US/08815809
2 Patent No. 6004777
3 GENERAL INFORMATION:
4 APPLICANT: Tartaglia, James
5 APPLICANT: Goebel, Scott J.
6 APPLICANT: Cox, William I.
7 APPLICANT: Gettig, Russell R.
8 APPLICANT: Pincus, Steven E.
9 APPLICANT: Pacletti, Enzo
10 APPLICANT: Jacobs, Hettam L.
11 TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF
12 FILE REFERENCE: 454310-1010
13 CURRENT APPLICATION NUMBER: US/08/815,809
14 CURRENT FILING DATE: 1997-03-12
15 NUMBER OF SEQ ID NOS: 23
16 SOFTWARE: PatentIn Ver. 2.0
17 SEQ ID NO 2
18 LENGTH: 2844
19 TYPE: DNA
20 ORGANISM: Vaccinia virus
21 US-08-815-809-2

alignment_scores:
1 Quality: 8.00 Length: 8
2 Ratio: 1.000 Caps: 0
3 Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
US-09-932-678-2 x US-08-815-809-2 ...

Align seq 1/1 to: US-08-815-809-2 from: 1 to: 2844
```

```

seq_name: /cpn2_6/ptdata/1/1na/6a_C988b.seq US-08-816-155B-4
seq_documentation_block:
Sequence 4, Application US/08816155B
Patent No. 5990091
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, JAMES
APPLICANT: COX, WILLIAM L.
APPLICANT: GELING, ROSSINI R.
APPLICANT: MARTINEZ, HECTOR
APPLICANT: PAOLETTI, ENZO
APPLICANT: PINOIS, STEVEN E.
TITLE OF INVENTION: METHODS HAVING ENHANCED EXPRESSION AND
NUMBER OF SEQUENCE: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FRIMMER, LAWRENCE A HAUG LLP
STREET: 745 FIFTH AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10151
C. COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
CREATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patcom Release #1.0, Version #1.10
DIFFERENT APPLICATION DATA:
APPLICATION NUMBER: US/08/816.155B
FILING DATE: 12 MAR 1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: KOMALSKI, THOMAS J.
REGISTRATION NUMBER: 42,147
REFERENCE/DOCKET NUMBER: 45410-2990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2856 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-816-155B-4
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Caps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_clock:
US-09-942-678-2 x US-08-816-155B-4
Align Seq 1/1 to: US-08-816-155B-4 from: 1 to: 2856
541 AcgSetThaLactYglYAspser 548
|||||
2596 CCGTCACACGGCTGGCTGGCATACG 2607
|||||
seq_name: /cpn2_6/ptdata/1/1na/6a_C988b.seq US-09-079-587-4
seq_documentation_block:
Sequence 4, Application US/09079587
Patent No. 6140066
GENERAL INFORMATION:

```



CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/012,688  
 FILING DATE: 03-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lavin J.C., Lawrence M.  
 REGISTRATION NUMBER: 30,768  
 REFERENCE/DOCKET NUMBER: 38-21(13583)A  
 TELEPHONE: (314)537-6670  
 TELEFAX: (314)537-6047  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3901 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-617-454-1

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-932-678-2 x US-08-617-454-1 ..  
 Align seq 1/1 to: US-08-617-454-1 from: 1 to: 3901

582 Gtuaspmetseralaglugluleu 589  
 |||||||  
 2293 GAGGATATGACCGTAGAAGACTG 2316

seq\_name: /cgen2\_6/prodata/1/lna/ECTUS\_COMB.seq: PCT-US94-01144-1

seq\_documentation\_block:  
 Sequence 1, Application PCT/US9401144  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: Plants Resistant to Infection by FLRV  
 NUMBER OF SEQUENCES: 5  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/01144  
 FILING DATE: 01-FEB-1994  
 CLASSIFICATION:  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3901 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA (genomic)  
 PCT-US94-01144-1

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-932-678-2 x PCT-US94-01144-1 ..  
 Align seq 1/1 to: PCT-US94-01144-1 from: 1 to: 3901

582 Gtuaspmetseralaglugluleu 589

|||||  
 2293 GAGGATATGACCGTAGAAGACTG 2316

seq\_name: /cgen2\_6/prodata/1/lna/6A\_COMB.seq: US 08-815-809-4  
 seq\_documentation\_block:  
 Sequence 3, Application US/08815809  
 Patent No. 6004777

GENERAL INFORMATION:  
 APPLICANT: TARIAGLIA, James  
 APPLICANT: GOEBEL, Scott J.  
 APPLICANT: COX, William L.  
 APPLICANT: GETTIS, Russell R.  
 APPLICANT: PINCUS, Steven E.  
 APPLICANT: PAOTELLI, Enzo  
 TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF  
 TITLE OF INVENTION: MAKING AND USES THEREOF  
 PRIORITY REFERENCE: 454,410-3010  
 CURRENT APPLICATION NUMBER: US/08/815,809  
 CORRESPONDING FILING DATE: 1997-03-12  
 NUMBER OF SEQ ID NOS: 23  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 3  
 LENGTH: 6628  
 TYPE: DNA  
 ORGANISM: Vaccinia virus  
 US-08-815-809-4

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-932-678-2 x US-08-815-809-4 ..  
 Align seq 1/1 to: reverse of: US-08-815-809-4 from: 1 to: 6628

541 Argserthraldeylglyaspser 548  
 |||||||  
 1251 GACTGACGCGCTGATGCGCAIAGC 1228

seq\_name: /cgen2\_6/prodata/1/lna/5H\_COMB.seq: US-08-816-1558-5

seq\_documentation\_block:  
 Sequence 5, Application US/088161558  
 Patent No. 5996091  
 GENERAL INFORMATION:  
 APPLICANT: TARIAGLIA, JAMES  
 APPLICANT: COX, WILLIAM L.  
 APPLICANT: GETTIS, RUSSELL R.  
 APPLICANT: MARTINEZ, HECTOR  
 APPLICANT: PAOTELLI, ENZO  
 APPLICANT: PINCUS, STEVEN E.  
 TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
 TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FROMMER LAWRENCE & HANG LLP  
 STREET: 745 FIFTH AVENUE  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: USA  
 ZIP: 10151  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/816,1558

FILING DATE: 12 MAR 1997  
 CLASSIFICATION: 514  
 ALTERNET/AGENT INFORMATION:  
 NAME: KOWALSKI, THOMAS J.  
 IDENTIFICATION NUMBER: 42,147  
 REFERENCE/DOCKET NUMBER: 45410-2990  
 TELEPHONE: 212-588-0800  
 TELEFAX: 212-588-0500  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6649 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US OR 016 1558 %

alignment scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Identity: 100.000 Percent Identity: 100.000

alignment block:  
 US-09-932-678-2 X US-08-016-1558 %rev

Align seq 1/1 to reverse of: US-08-016-1558-5 from: 1 to: 6649

541 A:GCGTGAAGTAACTGATGAGAGG 548  
 1272 CAGTCAAGGAGGATGAGATAGG 1249

Seq name: /seq02-6/ProdData/1/Inv/6A-0808.seq:US-09-079-587-5

Seq description block:

Sequence: 5, Application: US/09079587  
 Patent No.: 6150066  
 GENERAL INFORMATION:  
 APPLICANT: FARMACIA, JAMES  
 APPLICANT: GSK, WILLIAM J.  
 APPLICANT: GRIFFIN, RUSSELL R.  
 APPLICANT: MARTINEZ, DEYOUNG  
 APPLICANT: PAVETTI, EMZO  
 APPLICANT: JUNGUS, STEVEN F.  
 TITLE OF INVENTION: METHODS HAVING ENHANCED EXPRESSION, AND  
 TITLE OF INVENTION: METHODS OF MAKING AND USING THEREOF  
 NUMBER OF SEQUENCES: 48  
 ADDRESS/AGENT ADDRESS:  
 ADDRESS: 745 FIFTH AVENUE  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: USA  
 ZIP: 10151  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM pc compatible  
 OPERATING SYSTEM: PC DOS/MS DOS  
 SOFTWARE: Patented Release #1.0, Version #1.10  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/079,587  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US-08/016,155  
 FILING DATE: 12 MAR 1997  
 ALTERNET/AGENT INFORMATION:  
 NAME: KOWALSKI, THOMAS J.  
 IDENTIFICATION NUMBER: 42,147  
 REFERENCE/DOCKET NUMBER: 45410-2990  
 TELEPHONE: 212-588-0800

TELEFAX: 212-588-0500  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6649 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-09-079-587-5

alignment scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Identity: 100.000 Percent Identity: 100.000

alignment block:  
 US-09-932-678-2 X US-09-079-587-5/rev

Align seq 1/1 to reverse of: US-09-079-587-5 from: 1 to: 6649

541 A:GCGTGAAGTAACTGATGAGAGG 548  
 1272 CAGTCAAGGAGGATGAGATAGG 1249

Seq name: /seq02-6/ProdData/1/Inv/6A-0808.seq:US-08-244-478A-40

Seq description block:

Sequence: 40, Application: US/08244-478A  
 Patent No.: 5756423  
 GENERAL INFORMATION:  
 APPLICANT: Kallenberg, Socha  
 APPLICANT: Bouillon, Francois  
 TITLE OF INVENTION: METHOD FOR GENERATING STRUCTURAL AND  
 TITLE OF INVENTION: FUNCTIONAL DIVERSITY IN A DEPTIDE SEQUENCE  
 NUMBER OF SEQUENCES: 62  
 ADDRESS/AGENT ADDRESS:  
 ADDRESS: 1755 JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: USA  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM pc compatible  
 OPERATING SYSTEM: PC DOS/MS DOS  
 SOFTWARE: Patented Release #1.0, Version #1.10  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/244,478A  
 FILING DATE: 01-SEP-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 91/5489  
 FILING DATE: 11-SEP-1991  
 ALTERNET/AGENT INFORMATION:  
 NAME: GOLLION, NO. 5756423/40 F  
 IDENTIFICATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 660-078-0-FOCT  
 TELEPHONE: 703-413-4000  
 TELEFAX: 703-413-2220  
 INFORMATION FOR SEQ ID NO: 40:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 46 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-244-478A-40



alignment\_scores:  
 Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-932-678-2 x US-08-244-378A-40/rev

Align seq 1/1 to reverse of: US-08-244-378A-40 from: 1 to: 46

541 ArgSerThrAlaGlyGlyAsp 547  
 |||||  
 39 AcGTGCGCGCGGGGGGGGAT 19

seq\_name: /cqn2\_6/prodata/1/lna/5A\_COMB.seq:US-08-049-473-19

seq\_documentation\_block:

Sequence ID: Application US/08049473  
 Patent No: 5486021  
 GENERAL INFORMATION:  
 APPLICANT: Moss, Joel  
 APPLICANT: Mishima, Koichi  
 APPLICANT: Nightingale, Maria  
 APPLICANT: Tsuchiya, Mikako  
 TITLE OF INVENTION: A MAMMALIAN GUANIN NUCLEOTIDE BINDING  
 TITLE OF INVENTION: PROTEIN WITH AN ADP-RIBOSYLATION FACTOR DOMAIN  
 NUMBER OF SEQUENCES: 34  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: KNABBE, MARLENE, OLSON AND BEAR  
 STREET: 620 NEWPORT CENTER DRIVE SIXTEENTH FLOOR  
 CITY: NEWPORT BEACH  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92660  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/049,473  
 FILING DATE: 19940419  
 CLASSIFICATION: 436  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fuller, Michael L.  
 REGISTRATION NUMBER: 36,516  
 REFERENCE/DOCKET NUMBER: NIH050.001CP1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619-235-8550  
 TELEFAX: 619-235-0176  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 48 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-049-473-19

alignment\_scores:  
 Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-932-678-2 x US-08-049-473-19/rev

Align seq 1/1 to reverse of: US-08-049-473-19 from: 1 to: 48

176 SerAspSerAspAspGluAsp 176  
 |||||  
 45 TGGAGTCGAGCGATGAGGAT 19



Sequence	Strid	Orig	ZScore	Escape	Len	Documentation
qb.esl1:AM408066	+	171.00	3107.65	7.5e-164	563	AM408066 UT-HF-HMO-adv-b-10-0-0
qb.esl2:AC628405	+	169.00	3065.79	1.6e-161	1037	B64248305 D620498948F81 NIH MGC 7
qb.esl1:AC597171	+	167.00	3034.56	6.9e-160	550	ALS597171 DKR2313K0411 x1 313 6
qb.esl1:AA1213789	+	166.00	3016.47	9.0e-159	535	AA113789 chr9p11.81 NCL CGAP 99
qb.esl2:AB5615523	+	157.00	2848.87	2.0e-149	733	B5615523 60256262221 NIH MGC 99
qb.esl2:BE549643	+	152.00	2761.26	1.5e-144	458	BE549643 70b20908 x1 NCL CGAP 100
qb.esl1:AM548173	+	147.00	2667.87	2.4e-139	564	AM548173 EST370243 MAGE ressequ
qb.esl1:AA611628	+	146.00	2651.66	1.9e-138	440	AA611628 02b74d03 x1 NCL CGAP 100
qb.esl2:BA6170671	+	141.00	2553.12	5.8e-133	990	BA6170671 602244740401 NIH MGC 8
qb.esl1:BE0930377	+	140.00	2542.07	2.4e-132	422	BE0930377 RCS-BT0744-b-260400-031
qb.esl2:BE6611365	+	139.00	2519.47	4.3e-131	697	BE611364 60261272251 NIH MGC 61
qb.esl1:AA4812964	+	126.00	2286.44	4.1e-118	380	AA481295 U292605 x1 NCL CGAP 100
qb.esl1:AA491111	+	126.00	2285.56	4.6e-118	421	AA911111 zp86403 x1 STRataGene
qb.esl2:BE502966	+	125.00	2283.94	5.7e-117	996	BE502966 hb81907 x1 NCL CGAP 100
qb.esl2:BE5023864	+	122.00	2204.91	1.4e-113	996	BE5023864 hb18666331F1 NIH MGC 1
qb.esl1:AM703267	+	120.00	2174.52	7.1e-112	479	AM703267 AB38b04 x1 NCL CGAP 100
qb.esl1:AM703379	+	114.00	2061.74	1.4e-110	671	AM703379 AD8 Hmoa sap100
qb.esl2:BC149493	+	112.00	2027.59	1.1e-103	748	BC149493 nad29504 x1 NCL CGAP 6
qb.esl2:BC0502871	+	109.00	1968.70	2.1e-100	788	BC0502871 6025650403 NIH MGC 61
qb.esl2:BE614966	+	106.00	1915.62	1.9e-97	632	BE614966 6015048934F1 NIH MGC 7
qb.esl1:AM705483	+	103.00	1865.74	1.1e-94	348	AM705483 U1-H-B10-aag-f1-10-0-0
qb.esl2:AM768543	+	93.00	1679.29	2.7e-84	507	AM768543 hb62102 x1 NCL CGAP 100
qb.esl2:BA6149669	+	92.00	1665.05	1.7e-83	316	BA6149669 nad32041 x1 NCL CGAP 100
qb.esl1:AM4978208	+	91.00	1645.06	2.2e-82	382	AM4978208 AV639508 G1C Hmoa sap100
qb.esl2:BI6012138	+	88.00	1583.41	6.0e-79	835	BI6012138 6032457555F1 NIH MGC 99
qb.esl2:BA1646471	+	86.00	1547.84	5.7e-77	736	BA1646471 60230234241F1 NIH MGC 99
qb.esl2:BE0501864	+	80.00	1437.26	8.3e-71	793	BE0501864 6025448963F1 NIH MGC 61
qb.esl1:AM7709268	+	75.00	1351.61	4.9e-66	394	AM770928 h193e04 x1 NCL CGAP 100
qb.esl1:AI928274	+	75.00	1350.27	5.8e-66	461	AI928274 w067c08 x1 NCL CGAP 100
qb.esl2:BE621444	+	75.00	1349.63	6.3e-66	497	BE621444 7066912 x1 NCL CGAP 100
qb.esl2:BE110176	+	74.00	1349.61	6.3e-66	498	BE110176 7050605 x1 NCL CGAP 100
qb.esl1:AM978265	+	74.00	1332.06	6.0e-65	455	AM978265 EST390374 MAGE ressequ
qb.esl2:BE151543	+	73.00	1317.96	3.7e-64	278	BE151543 6018B08933F1 NIH MGC 5
qb.esl2:NA4382	+	72.00	1300.26	3.7e-63	258	NA4382 AY151111 FL Soares molan
qb.esl1:AI799954	+	71.00	1276.23	7.5e-62	503	AI799954 w641d02 x1 NCL CGAP 100
qb.esl1:AM668124	+	71.00	1275.80	8.1e-62	529	AM668124 EST380319 MAGE ressequ
qb.esl2:BE870692	+	69.00	1272.91	1.2e-61	742	BE870692 6014485286F1 NIH MGC 6
qb.esl1:AA213873	+	69.00	1240.57	7.5e-60	748	AA213873 chr9p11.81 NCL CGAP 99

TGGATAAACCAICITGACAAACCTCCT66 COTGATAAACAATCTG/

6' CTGATAACACATCTTGA



LOCUS AL597171 550 bp mRNA linear EST 14-AUG-2001  
 DEFINITION DKFZP313K0411.1 313 (synonym: hlec2) Homo sapiens cDNA clone  
 DKFZP313K0411.5', mRNA sequence.  
 ACCESSION AL597171  
 VERSION AL597171.1 GI:15154983  
 EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 550)  
 Koehler, K., Beyer, A., Mewes, W., Well, B. and Wiemann, S.  
 EST (Koehler, K., Beyer, A., Mewes, H.W., Well, B. and Wiemann, S.)  
 Unpublished (1999)  
 JOURNAL Contact: Koehler K  
 MIPs  
 Am Klepterspitz 16a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert.  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ): Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by BMEZ (Biomedical Research Center at the Charité,  
 Berlin/Germany) within the cDNA sequencing consortium of the German  
 Genome Project.  
 No SI sequence available.  
 This clone (DKFZP313K0411) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY. Email: clone@rzpd.de.  
 Location/Qualifiers  
 1..550  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFZP313K0411"  
 /clone\_1ib="313 (synonym: hlec2)"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Vector: pT7blue2; Site\_1: S11A; Site\_2: S11B;  
 cDNA-collection"  
 BASE COUNT 140 a 116 c 140 g 153 t 1 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 167.00 Length: 167  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 05-09-942-678-2 x AL597171 ..  
 Align seq 1/1 to: AL597171 from: 1 to: 550  
 1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaIaIaIaIa 17  
 48 AAGGAGGACAGCGCTGCTTCACAGCGCTTGGCCGACATCGCGCGCTTC 97  
 17 TGTSTSTAlaValLysLysLeuGlyAlaSerArgThrGlyIleSerAsnM 34  
 98 GTCTCTGTGAGTTAAGAACGTGGGGGTGAGGAGCTGAGATTCTTAATA 147  
 34 cTATGAlaLeuGluAsnAspPhePheAsnSerProPArgLysThrVal 50  
 148 TCGTGCCATTAGCAAGTACCTTTTCATCTCCGCCAAGAAAGCGCTG 197  
 51 ATGPhgGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGly 67  
 198 CGGTGTGGTGGACAGTGTGACAAAGCTGTGCTGACAGTACAAAGCGCTG 247  
 67 uThrAsnAspPheGluLeuLeuLysAsnGlnIleLeuLeuAspProAsp 44  
 248 AACAAATGACTTGTGAGTTGTGACAAACACCTGTGTGATTCACATATA 297  
 84 yAspAspPheGluLeuLeuLysAsnTrpLeuLeuGluThrArgSerSerIle 100  
 100

25c AATATGACCAATCATCAWCAACTGATGTATAGAAATTCCTTCTATCAAG 447  
 101 TTTTCTTTTAspPheGluLeuLeuLysSerIleIleLeuLeuLeu 117  
 34c TACTGTGACAAAGAACTTTCACCAACTATATAGATATATTAAGATTCT 397  
 117 CTTPLeuAsnArgSerGlnThrValValGluGluGlyTyrLeuAlaPheGlu 144  
 39c TTGCTTGAATGAACTCAAAACTAGTGAAGAGATTTTGATTTTCTTG 447  
 144 TAsnLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150  
 44c GATACCTGTATACCAACAGACAGCTGTTCCTTCAGACCGTGTTCACAG 497  
 151 TLeuAlaSerHisPheValProArgValIleIleLysGluGlyAspVal 167  
 49c ATTGCTTCCATTGTGTGTCTCCGAGTATCATTAAAGAAAGCAATGT 547  
 167 1 167  
 54c A 548  
 seq\_name: kb\_esi1:AA213789  
 seq\_documentation\_block:  
 LOCUS AA213789 548 bp mRNA linear EST 14-AUG-1997  
 DEFINITION Z791411.s1 NCL\_CvAF\_GCB1 Homo sapiens cDNA clone IMAGE:684064.  
 mRNA sequence.  
 ACCESSION AA213789  
 VERSION AA213789.1 GI:1812416  
 KEYWORDS EST.  
 SOURCE Human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 548)  
 NCL-CvAF <http://www.ncbi.nlm.nih.gov/ncicvaf>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CvAF).  
 TITL Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: rcpu@remail.nih.gov  
 This clone is available royally-free through the IMAGE Consortium (<http://image.dlbb.gov>) for further information.  
 Possible reversed clones: polyt not found  
 Insert length: 1076 Std Error: 0.00  
 Seq primer: 41m13 fwd: ET from Amersham  
 High quality sequence stop: 457.  
 Location/Qualifiers  
 1..548  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:684064"  
 /clone\_1ib="NCL\_CvAF\_GCB1"  
 /issue\_type="germline center B cell"  
 /lab\_host="DH10B"  
 /note="Vector: pT7-td Pac (pharmacia) with a modified  
 polylinker. Site\_1: Not 1; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (Ch21c, 14p7).  
 provided by Dr. Louis M. Staudt (NCI). Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CHBR). cDNA synthesis was  
 primed with a Not I oligo(dT) primer  
 15'-TCTTACCAATCTGAACTGGACAGGCGCTTATTTTCTTTTCTT 4'  
 1. Double-stranded cDNA was ligated to Eco RI adaptors  
 (pharmacia). digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Benito Soares and M. Fatima Brualdo."  
 BASE COUNT 150 a 102 c 121 g 161 t 1 others  
 ORIGIN



151 *HeLa*SerHisPheValPro 152  
 |||||  
 474 ATGGCTCCCATTTTGTGGCT 499





seq\_documentation\_block: 990 bp mRNA linear EST 06-FEB-2001  
 LOCUS BC167061  
 DEFINITION 60234470F1 NIH\_MGC-89 Homo sapiens cDNA clone IMAGE:4455009.5,  
 mRNA sequence.  
 ACCESSION BC167061  
 VERSION BC167061.1 GI:12673764  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 990)  
 NIH-MGC <http://imgc.cci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [raab@t-mail.nih.gov](mailto:raab@t-mail.nih.gov)  
 Tissue procurement: ATCC  
 cDNA library preparation: Life Technologies, Inc.  
 cDNA library Arrayed by: The I.M.A.G.E. Consortium (ILINC)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILINC at:  
<http://image.llnl.gov>  
 Plate: L1AM10247 row: 0 column: 10  
 High quality sequence stop: 665.  
 Location/Qualifiers  
 1..990  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NIH\_MGC\_89"  
 /clone\_11b="NIH\_MGC\_89"  
 /tissue\_type="hypertrophic, cell line"  
 /lab\_host="BHI10B (phage-resistant)"  
 /note="Organ: Kidney; Vector: PCMV-SPOK16; Site: 1; Note:  
 Site 2: Tail: Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.3 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: This is a NIH-MGC library."  
 BASF COUNT 304 a 275 c 219 g 192 t  
 ORIGIN  
 alignment\_scores: Quality: 141.00 Length: 141  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-932-678-2 x BC167061  
 Align seq 1/1 to: BC167061 from: 1 to: 990  
 501 Proluyslllecyllouproscrvalvalasrphthalalaleth 517  
 |||||||  
 108 CCCCTGAGAGATTTCCTGCGCTCAGTGTTAACTTTTTCCTCATAC 157  
 |||||||  
 517 TASNLYSTYRGInleuValPheCYSTYRThrllleGluAGAsAsna 534  
 |||||||  
 158 AATATACACACCGCTGCTGCTGACACCATTTGAGAGACAAATC 207  
 |||||||  
 534 TGGTlmePleuProvalleargSerThrAlaGlylAspserValGln 550  
 |||||||  
 208 GCGCATCTGCGCGAGTCATTAGAGAGTAGCGCTGAGAGACAGACG 257  
 |||||||  
 551 llecylthAsnProleuAspThrPhePheProPheAspProcyValle 567  
 |||||||  
 258 ATCTGCAACAACCGCTGAGACCTTCTCCCTTTGATGCCGCTGCT 307  
 |||||||  
 567 ATGATGSerlAsylAspPheleAspProleuArglnValTPGluAspM 584  
 |||||||  
 308 CAAGAGGTCAGAAATTCATTATCTATTATCAAGAGTGAGAGACA 357

seq\_documentation\_block: 422 bp mRNA linear EST 12-JUN-2001  
 LOCUS BC094077  
 DEFINITION KC5-B10744-260400-031 H10 B10744 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BC094077  
 VERSION BC094077.1 GI:8484529  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 422)  
 Dias, M.C.R., Garcia-Correa, K., Wojcowski-Almeida, S., Brites, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zaqui, M.A., Bordin, S., Costa, P. Jr.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Hata, G.S., Simpson, J.H.,  
 Rutschkin, A., deoliveira, P.S., Bucher, P., Jomparelli, V., O'Hare,  
 M.J., Soares, F., Brenlund, R.R., Reis, L.F., de Souza, S.J., and  
 Simpson, A.J.J.  
 TITLE Shotgun sequencing of the human transcriptome with over-expressed  
 sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.J.,  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,  
 Brazil  
 TEL: +55-11-2704422  
 FAX: +55-11-2707001  
 Email: [astimpson@ludwig.org.br](mailto:astimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/110R Human Cancer Genome  
 Project. This entry can be seen in the following DBI  
 (<http://www.ludwig.org.br/scripts/setlm12.pl?L1412-B05-B10744-260>  
 400-031-B10744-2000-04-26a14-1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 414.  
 Location/Qualifiers  
 1..422  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="B10744"  
 /seq\_stage="Adult"  
 /note="Organ: breast; Vector: puc18; Site: 1; Site 2:  
 Site 3: A mini-library was made by cloning products derived  
 from ONESIES PCR (U.S. Letters Patent application No. 196  
 7716 - Ludwig Institute for Cancer Research) products  
 into the pUC 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."  
 BASF COUNT 121 a 98 c 99 g 104 t  
 ORIGIN  
 alignment\_scores: Quality: 140.00 Length: 140







**DEFINITION** 60186633f1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:409398 5', mRNA sequence.  
**ACCESSION** HF203684  
**VERSION** HF203684.1 GI:11097270  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
**REFERENCE** 1 (bases 1 to 996)  
**AUTHORS** NIH-MGC <http://mgs.sci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: [rgs@nhi.nih.gov](mailto:rgs@nhi.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The J.M.A.G.E. Consortium (LJNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the J.M.A.G.E. Consortium/LJNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LJCW67 row: n column: 07  
 High quality sequence stop: 684.  
 Location/Qualifiers  
 1..996  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:409398"  
 /clone\_1lb="NIH\_MGC\_17"  
 /tissue\_type="chondrosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: muscle; Vector: pOT8; Site\_1: EcoRI; Site\_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
**BASE COUNT** 318 a 211 c 231 g 216 t  
**ORIGIN**

**alignment\_scores:**  
 Quality: 122.00 Length: 122  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

**alignment\_block:**  
 US-09-932-678-2 x HF203684 ...

Align seg 1/1 to: HF203684 from: 1 to: 996

```

430 ValAsnTrpLeuHisIleTyrLeuAsnAsnGlnAspSerGlyThrIysAl 446
|||||
155 GTTAACCTGGCTGCACATATACCTTATATACGACGATTGGGAAACAAAGG 204
|||||
446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysGlnAlaVal 463
|||||
205 ATTTCGAGAGTGTGCTTCATGAGCATTTTACTCAGCTGGCAACCTG 254
|||||
463 aPheTyrThrPheValPheArgHisIleGlyIleLeuLeuSerGlyAsnLeu 479
|||||
255 TGTTTACACCTTTGTTTATACACGACGACCTTTTGACCGCAAAACCTG 304
|||||
480 LysGlnGlyLeuGlnTyrLeuGlnSerLeuAsnPheGluArgIleValMet 496
|||||
305 AAAGAGAGGTTTTCAGATATCTCAGAGCTCGAATTTGACGCGATATGAT 354
|||||
496 tSerGlnLeuAsnProIlePheIleCysLeuProSerValValAsnPheP 513
|||||
355 GAGCCAGGCTAATCCCTGAGATTGACTGCCCTCAGTGGTTACTTTT 404

```

```

513 LeuAlaAlaIleThrAsnLysIleGlnIleValIlePheCysTyrPheIleIle 529
|||||
405 TTGCTGCATCATCAAAATATAGTACGTCTGCTTCTGTACACATCATCAT 454
|||||
530 GUAATGAsnAspArgIleMetLeuProValIleArgSerThrAlaIleVal 546
|||||
455 GAGAGGAAACAATCGCCAGATGCTGCTCAGTATTAGAGATACCTCTGGAG 504
|||||
546 yAspSerValGlnIle 551
|||||
505 AGACTCAGCTCAGATC 520

```









```

|||||
1919 TACCGTTCATGGGATACACCAAGCTCTTGACAGCATTTCCCAATTC 1968
634 rosetSerValGlySerProValLeuTyrMetGlnProSerPro 650
|||||
1969 CTCACAGTAGTGGGCTCCGCCACCGCTGTCACATGCACCGCATGCC 2018
651 leu 651
|||||
2019 CTC 2021

seq_name: qb-pr: HSA272050
seq_documentation_block:
LOCUS HSA272050 3756 bp mRNA linear PRI 07-SEP-2000
DEFINITION Homo sapiens mRNA for transcription initiation factor 1A protein
(TIF-1A gene).
ACCESSION AJ272050.1 GI:10046713
VERSION AJ272050.1
KEYWORDS TIF-1A; transcription initiation factor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3756)
AUTHORS Boden, J., Hoffmann-Kohrer, U., Koss, W., Deltus, H., Vingron, M. and
Grunml, I.
TITLE Cloning and functional characterization of transcription initiation
factor TIF-1A, a growth-dependent regulator of ribosomal RNA
synthesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3756)
AUTHORS Boden, J.J.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2000) Boden J.J., Molekulare Biologie der Zelle
11, Deutsches Krebsforschungszentrum, INF 280, 69120 Heidelberg,
GERMANY

FEATURES
source location/qualifiers
1..3756
/organism="Homo sapiens"
/gb_xref="taxon:9606"
23..3756
/gene="TIF-1A"
23..3756
/gene="TIF-1A"
/function="transcription initiation factor"
/codon_start=1
/evidence=experimental
/product="transcription initiation factor 1A protein"
/protein_id="CA07955.1"
/gb_xref="GI:10046713"
/translat="MAAPLHLRLPGDAASSAVKLGASKTGLSNMRLKLENDPENS
PRLKTVRGVGTVEVLKYEKGEINDELKNOLOPDIKDOILMLLEFSSIMYL
TKDFEOLISILRLPMLRSQVEEYLAFLGSLVAGQVYPLPCISMLASIEPVR
LKEGVIVSDSDDEDNLPANFDICRALOLLIARVPSIPMLMGLIYEKPEYKRS
ERTCEYVHNLRISTVPTLRHLELLEIKLIDVNASRGLIDAEATATOTCG
TUSTEGLNMDDEETETKAPRELDQVHYVAERLILNSLYSTKDYCKYDK
VINGRTKDLIRDLINIFDKLPLTHASCHVQFTPTVCSKUDFAFARLEHLKRLD
PSPALIRQAGANTIGSFLARAFPLIIVKSLDLVWLMLHYLNDSGTRKADIV
ALHGPFYSAGCAVPEYVVERKRLSGNLKEGLYQSLNFRIVMSQNPILCLPS
VNVFPAALTNKYQLFCYTLIERNNRQMLPVKSTFGSGSVQICRLDTEFPDPCV
LKRSKKEIDPIYQVWEDMSAEHIDFKPKPKKIDVEIDHDFIKRGVQNDVIGITP
SSEDTHPSPSSSVSGPVLVMSPL"
3756
/gene="TIF-1A"
/evidence=experimental

BASE COUNT 1060 a 692 c 844 g 1156 t 4 others
ORIGIN

alignment_scores:
Quality: 550.00 length: 651
Ratio: 0.846 Gaps: 0

```

```

Percent Similarity: 99.846 Percent Identity: 99.846
alignment_block:
US-09-932-678-2 x HSA272050
Align seq 1/1 to: HSA272050 from: 1 to: 3756

1 MetAlaAlaProLeuLeuLeuThrArgGluPheGluYAspAlaAlaIleIle 17
|||||
21 AAGGGGAGACCTGGCTTCACAGCGCTTGGGGAGATGGCGGGCTG 72
|||||
17 T SerSerAlaValIstValLeuGlyValMetArgThrGlyIleSerAsn 44
|||||
73 GTCCCTGCATTAAGAAAGCTGGCGCGTCGAGAGCTGGATTCAAATA 122
|||||
34 eValGAlaIleLeuIAspAspPhePheAsnSerProProAlaIstVal 50
|||||
128 TGGGTCATACGAAAGACTTTTCAATTCTGCCCAAGAAAGAGCT 172
|||||
51 ArgPheGlyGlyThrValPheLeuValLeuLeuIstValIstValGly 67
|||||
17 CGATTGGTGTAAGCTGTGACATAGAGCTTGTGAGAGTCAAAAAGCGTGA 222
|||||
47 GlnAspAspPheGluLeuLeuIstValIstValIstValIstValIst 84
|||||
227 AACCAATGACCTTGAGTGTGAGAGAGCGAGCTGTACATCCACACATAA 272
|||||
84 YAspAspAlaIleLeuAsnTrpLeuLeuGluPheArgSerSerIleMet 300
|||||
277 AACAGACACAGATTCATCAATGAGCTGTAGAAATTCGCTGTCTATACAG 322
|||||
141 TTrpGluThrIstAspPheGluGluGluIstSerIleLeuValLeuP 117
|||||
429 TACTTGCAAAAGACCTTCAGTACATATACATATATATATATATTCCT 372
|||||
117 GTPLPLeuAsnArgSerGlnThrValValGluGluIstValLeuAlaPheGlu 134
|||||
373 TTGGTGCATACAAAGCAAAAGCACTAGTCGACAGCATTTGGCTTTTGG 422
|||||
114 YAsnLeuValSerAlaGlnThrValPheLeuArgProGlySerMet 150
|||||
427 GAACTGTGATGCAACACAGATGTTGTTCACAGCGCTGTTCAGATG 472
|||||
151 IleAlaSerHisPheValProProArgValIleIleIstValIstVal 167
|||||
477 ATTGGTTCCTCATTTTGGCTGTCTGAGATGATTAAGAAAGAGATG 522
|||||
167 IAspValIstAspSerAspAspAlaIstAspAspAlaLeuProAlaAsnHis 184
|||||
527 AACATGTTCACATTTGCAATGATACAGATGATATCTCTCTGAAATTTTG 572
|||||
184 STrpTrpSerHisSerAlaAlaLeuGluIleIleAlaArgValValIstSer 200
|||||
573 ACACATGTGCAATAAGCTTTCACATATATAGTAAAGATAGTACACATGCA 622
|||||
201 PTrpTrpPheLeuMetTrpGluLeuValGluIstValIstValIstVal 217
|||||
623 CGTGGCTTTCATGCGCAATATCGTGGGAAATTTTCATTTGTTCAAA 672
|||||
217 SerGluArgThrLeuGluIstValIstValIstValIstValIstVal 234
|||||
672 AAcAGACAGAAAGATACAAAGTACGTCATTAAGTAAAGATATAGT 722
|||||
234 AlTrpPheProThrTrpArgHisGlnIleLeuGluLeuIleLeuLeu 250
|||||
723 TATATTTCCAATTTGAGGATGAATTTGAGCTTATTTATTAATAAA 772
|||||
251 LeuLeuIstValAspValAlaAlaIstArgGlnGlyIleGluIstAlaG 267
|||||
773 CTACTCAAGTTGAGTGTGAATGCAATCCGCGAGGATATGAAGATCTGA 822
|||||
267 GlnThrAlaThrGluThrGlyGlyThrAspSerThrGluGluLeuP 284
|||||

```

[illegible]





```

234  aattpppctrothleatgthgllcgtttlelengllleuplletl)ctolnlys 250
361  tatattttccacatttgatgacgcatacaattctgacctattattatgaaaa 430

251  leuleulstleuaspya/asnalastatgctngtlyllcgttuaspaalac1 267
|||||
441  ctactcgaagcttgcatgtgatgcattcccgccagcggatatttgaaacatcctga 480

267  uclatthra1a 270
|||||
481  acgaacacaca 490

seq_name: qb_pr:BC009198

seq_documentation_block:
LOCUS      BC009198              1177 bp      mRNA      linear      12-JUL-2001
DEFINITION Homo sapiens, similar to RNA polymerase I transcription factor
ACCESSION  BC009198
VERSION    BC009198.1  GI:14327947
KEYWORDS   MGC.
SOURCE      human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eultheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1177)
AUTHORS   Strausberg/R
TITLE      Direct Submission
COMMENT    Submitted (06-JUN-2001) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA
REMARK     NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT    Contact: MGC help desk
           Email: mgc-help@mail.nih.gov
           Tissue procurement: Louis M. Staudt, M.D., Ph.D.
           cDNA Library Preparation: Rubin Laboratory
           DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Institute for Systems Biology
           http://www.systemsbio.org
           Contact: amadan@systemsbio.org
           Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
           Greene, Mark Kelleman and Anuradha Madan

clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: LNL Plate: 24 Row: K Column: 14
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA q1: 335302.
Location/Qualifiers
1..1177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:15121 IMAGE:3678732"
/tissue_type="Lymph, Burkitt lymphoma"
/clone_lib="NHL_MGC_8"
/lab_host="DH10B-R"
/note="Vector: pOH17"
316..636
/codon_start=1
/product="Similar to RNA polymerase I transcription factor
kRNY"
/protein_id="AAH09198.1"
/db_xref="GI:14327948"
/translation="MAHALNDPNSPPKTVRGSTVTVLTKYKGETNDPELKNQ
LTPADPKDDIIMTLLEFRSSVMYLTGDFOLLISILRLFWKRSQVVEVYLAIGRN
LVA*"
BASE COUNT      339 a      222 c      253 g      363 t
ORIGIN

```

```

Percent      Similarity: 100.000      Percent Identity: 100.000
Attachment_block:
US-09-942-678-2 x BT009198
Align seq 1/1 to: bc009198 from: 1 to: 1177
31 11cSerAsmMetArGa1aJaSuG1uInAsnAspPhoPheAsnSerPro1Gat 47
|||||
307 ATTCAATATATGCTGATTTAGAGATGATTTTTCATATCTGCTCCAAAG 456
47 glySThrValArgPheGlyGlyThrValThrGluValLeuLeuLysVal 64
|||||
357 AAAAAGCTTCCGCTTGCTGCACTGTGACAGAGCTCTGCAAGATCA 406
64 YSTSGSYGCUInhrAsnAspPheGluLeuLeuLysAsnGluLeuLeuGlu 80
|||||
407 AAAAGGCTGAACCAATGACTTGATGATTTGTTCAAGAAAGCTGTACAT 456
61 ProAspTTCysAspAspGlnTelleAsnTThrLeuLeuGluPheArg 97
|||||
457 CCGAGCAATAAAGATGACCAATCAATCAATCTGCTGCTAGAAATTCGTT 506
97 rSet 98
|||||
567 TTTCT 510
seq_name: j0_bhg:AC040158
seq_documentation_block:
LOCUS AC040158 129473 bp DNA linear HTG: 04-SEP-2000
DEFINITION Homo sapiens chromosome 16 clone CTA-1344, WORKING DRAFT SEQUENCE.
14 unordered pieces.
AC040158 AC040158.4 G1:96655A6
HTG: HTGS_PHASE1: HTGS_DRAFT:
SOURCE human:
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eulioria; Primates; Catarrhini; Hominoidea; Homo.
DOE Joint Genome Institute.
Sequencing 91 Human Chromosome 16
2 (bases 1 to 129473)
Unpublished
DOE Joint Genome Institute.
Direct Submission
Submitted (11-Apr-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
on Sep 3, 2000 this sequence version replaced g1:909085f2.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Project Name: 0
Center clone name: C17978SKA_13B4
-----
Summary Statistics
Consensus quality: 99046 bases at least Q40
Consensus quality: 99266 bases at least Q30
Consensus quality: 104098 bases at least Q20
Estimated insert size: 125000; adarose-tp estimation
Estimated insert size: 124173; sum-of-contigs estimation
Quality coverage: 4.51 in Q20 bases; adarose-tp estimation
Quality coverage: 4.41 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a working draft sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```



\* 101721 122163: contig of 20443 bp in length  
 \* 122164 122263: gap of unknown length  
 \* 122264 143900: contig of 21637 bp in length.  
 FEATURES  
 source 1..143900  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome:"16"  
 /clone="RP11-344H15"  
 /clone\_lib="RPC1 human BAC library 11"  
 /origin="34128 c 34320 g 35886 t 1775 others

BASE COUNT 17791 a 34128 c 34320 g 35886 t 1775 others  
 ORIGIN

## alignment\_scores:

Quality: 62.00 Length: 62  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-932-678-2 x AC092137/rev ..

Align seq 1/1 to reverse of: AC092137 from: 1 to: 143900

421 Thrvallyssercysleuaspleulevalasntprlchisiletylle 437  
 |||||||  
 66649 ACTGAAATATATGCTAGATCTTTGGTTAACGCTGCACATATACCT 66600

447 uasnasndinsperserilythrylsalapheycysapvalalaleuhisg 454  
 |||||||  
 66599 TAATACAGGATTCGGACAAACGATTCGCGATCTGCTTCGCAATG 66550

454 lypropheetyseralacysgialavalphetylrthryvalphary 470  
 |||||||  
 66549 GACCATTTATACGCTCCAGCTGTGTTCACACCTTTGTTTTCAGA 66500

471 Hslsyglnleuleuserglyasnluleysglugly 482  
 |||||||  
 66499 CACACGACGCTTTGACGCGAACCTGAAAGAGCT 66464

seq\_name: gb\_hhg:AC007615

## seq\_documentation\_block:

LOCUS AC007615 175691 bp DNA linear HTG 03-JUL-2001  
 DEFINITION Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT  
 SEQUENCE 1 ordered pieces.  
 AC007615  
 AC007615 6 GI:14589428  
 HTG: HTGS\_PHASE2: HTGS\_DRAFT: HTGS\_ACTIVEPIN.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 175691)  
 DOE Joint Genome Institute.  
 Sequencing of Human Chromosome 16  
 Unpublished  
 2 (bases 1 to 175691)  
 Jones, D., Mundt, M., Doggett, N., Monk, C., Saunders, E., Robinson, D.,  
 Bruce, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
 Bryant, J., Tesmer, J., Neimke, L., Longmire, J., White, S., Tatum, O.,  
 Campbell, C., Fawcett, J., Malbelle, M., Bussod, M., Sutherland, R.,  
 McMurtry, K., Han, C. and Deaven, L.  
 Direct Submission  
 Submitted (20 MAY 1999) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M888, Los  
 Alamos, NM 87545, USA  
 on Jul 3, 2001 this sequence version replaced a:13928651.

COMMENT  
 TITLE  
 JOURNAL

Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than

1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.

Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.

\* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 175691: contig of 175691 bp in length  
 FEATURES  
 source 1..175691  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome:"16"  
 /clone="RP11-528K16"  
 /clone\_lib="RPC1 human BAC library 11"  
 /origin="42192 a 42125 c 45277 g 45896 t 1 others

BASE COUNT 42192 a 42125 c 45277 g 45896 t 1 others  
 ORIGIN

alignment\_scores:  
 Quality: 62.00 Length: 62  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-932-678-2 x AC007615 ..

Align seq 1/1 to: AC007615 from: 1 to: 175691

421 Thrvallyssercysleuaspleulevalasntprlchisiletylle 437  
 |||||||  
 31067 ACTGAAATATATGCTAGATCTTTGGTTAACGCTGCACATATACCT 31056

447 uasnasndinsperserilythrylsalapheycysapvalalaleuhisg 454  
 |||||||  
 31057 TAATACAGGATTCGGACAAACGATTCGCGATCTGCTTCGCAATG 31066

454 lypropheetyseralacysgialavalphetylrthryvalphary 470  
 |||||||  
 31107 GACCATTTATACGCTCCAGCTGTGTTCACACCTTTGTTTTCAGA 31156

471 Hslsyglnleuleuserglyasnluleysglugly 482  
 |||||||  
 31157 CACACGACGCTTTGACGCGAACCTGAAAGAGCT 31192

seq\_name: gb\_hhg:AC092562

## seq\_documentation\_block:

LOCUS AC092562 187443 bp DNA linear HTG 18 OCT 2001  
 DEFINITION Papio hamadryas clone RP41 285113, WORKING DRAFT SEQUENCE  
 SEQUENCE 1 ordered pieces.  
 AC092562  
 AC092562 3 GI:16258968  
 HTG: HTGS\_PHASE1: HTGS\_DRAFT: HTGS\_FULLDRAFT: HTGS\_ACTIVEPIN.  
 SOURCE  
 baboon.  
 ORGANISM  
 Papio hamadryas  
 Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;

















GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:40:21 ; Search time 1685.89 seconds  
(without alignments)  
16536.064 Million cell updates/sec

Title: US-09-932-678-1

Perfect score: 2058

Sequence: 1 acagagcgtctgtgtgaac.....tgaactttgatttcccat 2058

Scoring table:

Gapop 60.0 ; Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:  
1: cm\_estiba:\*  
2: cm\_estiba:\*  
3: cm\_estiba:\*  
4: cm\_estiba:\*  
5: cm\_estiba:\*  
6: cm\_estiba:\*  
7: cm\_estiba:\*  
8: cm\_estiba:\*  
9: cm\_estiba:\*  
10: cm\_estiba:\*  
11: cm\_estiba:\*  
12: cm\_estiba:\*  
13: cm\_estiba:\*  
14: cm\_estiba:\*  
15: cm\_estiba:\*  
16: cm\_estiba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	547	26.5	550	9	AI597171 DKEP313K
2	510	24.7	788	10	BG502871 602550430
3	500	24.2	535	9	AA213789
4	497	24.0	1037	10	BG428305
5	467	22.6	563	9	AA408066 01-BF-EMO
6	460	22.2	733	10	B1561523
7	458	22.1	458	10	BES49643
8	449	21.7	671	9	AV703279
9	444	21.5	564	9	AM958173
10	415	20.1	697	10	BG611364
11	389	18.8	440	9	AA811628
12	385	18.6	996	10	BF203684
13	380	18.4	380	9	AA481295
14	375	18.1	490	10	BE502966
15	371	18.0	990	10	BG167061
16	371	17.9	422	9	BE093077
17	357	17.3	479	9	AM239267

18	337	16.3	498	10	BG149493
19	312	15.1	348	9	AM015483
20	300	14.5	421	9	AA191111
21	283	13.7	507	9	AM768543
22	278	13.4	316	10	BG149669
23	273	13.2	382	9	AV649508
24	273	13.2	828	10	B1761160
25	268	13.0	503	9	A1799954
26	267	12.9	632	10	BH614095
27	264	12.8	497	10	HE221544
28	264	12.8	498	10	HE110176
29	261	12.6	736	10	H1464721
30	260	12.6	793	10	BG501864
31	257	12.4	835	10	B1601238
32	246	11.9	461	9	A1928274
33	232	11.2	370	9	B166014
34	218	10.5	258	10	N42382
35	217	10.5	742	10	HE870692
36	214	10.3	529	9	AM968124
37	209	10.1	448	9	AA213873
38	209	10.1	727	10	B1549394
39	205	9.9	946	10	BG535886
40	203	9.8	356	10	HE832086
41	197	9.5	253	9	AA376316
42	197	9.5	278	10	BF215433
43	196	9.5	406	10	B1018122
44	196	9.5	505	10	BG532530
45	193	9.3	327	9	AM802818

# ALIGNMENTS

RESULT 1  
AI597171  
LOCUS  
DEFINITION  
DKEP313K0411.1 571 bp mRNA linear EST 14 AUG-2001  
DKEP313K0411.1 571 bp mRNA sequence.

VERSION  
AI597171.1 GI:15154984  
KEYWORDS  
SOURCE  
ORGANISM

human.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
Koehler, K., Beyer, A., Mewes, W., Well, B. and Wiemann, S.  
EST (Koehler, K., Beyer, A., Mewes, W., Well, B. and Wiemann, S.)  
Unpublished (1999)

JOURNAL  
COMMENT  
Contact: Koehler K

Am Klopsterspitz 18a D-82152 Martinsried, Germany  
This is the 57 sequence of the clone insert

Research Center (DKEP313K0411) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubergweg 6, 10559  
Berlin-Charlottenburg, GERMANY; Email: clone-rzpd.de.

FEATURES  
location/qualifiers

1..550  
/organism "Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKEP313K0411"  
/clone\_lib="13 (synonym: hlec2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pT7phx2; Site\_1: S11A; Site\_2: S11B;  
cDNA-collection"

BASE COUNT 140 a 116 c 140 g 153 t 1 others





RESULT	3	535 bp	MRNA	linear	EST 13-MUG-1997
AA213789	AA213789	535 bp	MRNA	linear	EST 13-MUG-1997
LOCUS	2791411.s1 NC1:CCAP.GCB1	535 bp	MRNA	linear	EST 13-MUG-1997
DEFINITION	mRNA sequence.				
ACCESSION	AA213789				
VERSION	AA213789.1				
KEYWORDS	EST.				
SOURCE	Human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia: Primates: Catarrhini: Hominidae: Homo				
REFERENCE	1 (bases 1 to 535)				
REFERENCE	NC1:CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .				
REFERENCE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).				
REFERENCE	Tumor Gene Index				
REFERENCE	Unpublished (1997)				
REFERENCE	Contact: Robert Strausberg, Ph.D.				
REFERENCE	Email: ccaps-r@mail.nih.gov				
REFERENCE	This clone is available royalty-free through LIND: contact the				
REFERENCE	IMAGE Consortium (info:Image.LIN.gov) for further information.				
REFERENCE	Insert length: 1076 Std Error: 0.00				
REFERENCE	Seq primer: 41ml3 fwd. Et from Amersham				
REFERENCE	High quality sequence stop: 457.				
REFERENCE	Location/Qualifiers				
REFERENCE	1..535				
REFERENCE	/organism="Homo sapiens"				
REFERENCE	/db_xref="Gene:5586356"				
REFERENCE	/db_xref="taxon:9606"				
REFERENCE	/clone="IMAGE:683036"				
REFERENCE	/clone="NC1:CCAP.GCB1"				
REFERENCE	/library="type="germlinal center B cell"				
REFERENCE	/lab_host="DH10B"				
REFERENCE	/note="Vector: p1773-Pac (Pharmacia) with a modified				
REFERENCE	polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand				
REFERENCE	was prepared from human tonsillar cells enriched for				
REFERENCE	germlinal center B cells by flow sorting (CD20+, IgD-),				
REFERENCE	provided by Dr. Louis M. Staudt (NCI), Dr. David Allman				
REFERENCE	(NCI) and Dr. Gerald Martin (CEBR). cDNA synthesis was				
REFERENCE	primed with a Not I - oligo(dT) primer				
REFERENCE	15'-GGTACCAATCTCAAGTGGAGGCGGCGCTCATTTTCTTTTCTTTT-3'				
REFERENCE	I. Double-stranded cDNA was ligated to Eco RI adaptors				
REFERENCE	(Pharmacia), digested with Not I and cloned into the Not I				
REFERENCE	and Eco RI sites of the modified p1773 vector. Library				
REFERENCE	went through one round of normalization, and was				
REFERENCE	constructed by Bento Soares and M. Fatima Honaldo."				
BASE COUNT	150 a 102 c 121 g 161 t 1 others				
ORIGIN					
Query Match	24.28; Score 500; DB 9; Length 535;				
Best Local Similarity	100.0%; Pred. No. 1.3e-247;				
Matches 500; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
DB	7 GCCTTCCTCTCTGCACTTTAAGAAAGCTGGCGCTGTCGACACTGGCATTTTCAAAATATGCG 66				
DB	173 TCAATTAAGAAATGACTTTTCAATCTCTCCCAAGAAATCTGTGGTTGGTGGAC 212				
DB	67 TCAATTAAGAAATGACTTTTCAATCTCTCCCAAGAAATCTGTGGTTGGTGGAC 126				
DB	213 TCAATTAAGAAATGACTTTTCAATCTCTCCCAAGAAATCTGTGGTTGGTGGAC 292				
DB	127 TGTCAAGAAATGACTTTTCAATCTCTCCCAAGAAATCTGTGGTTGGTGGAC 186				
DB	293 GAATTAAGAAATGACTTTTCAATCTCTCCCAAGAAATCTGTGGTTGGTGGAC 352				
DB	187 GAATTAAGAAATGACTTTTCAATCTCTCCCAAGAAATCTGTGGTTGGTGGAC 246				

OY	353	ccctcctctacatgacctgaacaaagactttggagaattatcaatattatattaa	412
Dd	217	cccccttccttcatgataisnctgcacaagaagcttgtagaatgtatatcataattatag	406
OY	413	attacccttaattgaatgaagatcaaacagtatattcgaaagatattgaatttcttcaatga	472
Dd	367	attcccccttttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc	466
OY	473	tcttgtatctgaagcaagcgtttctctcaagcctgatctcagatatttgatctctctt	532
Dd	467	tcttgtatctgaagcaagcgtttctctcaagcctgatctcagatatttgatctctctt	426
OY	533	tatgctctcccagatattgattcattgaagaaaggagatgttatgtttcgaattctatattad	592
Dd	427	tggctctcccgacgattatcattaacgaaagcgaaagctatgatttcagattctgattatcca	486
OY	593	aagatgaatactctctctcaaa	612
Dd	487	agatgaataactctctctcaaa	506
RESULT	4		
BG42B305		MC42B305	1037 bp
LOCUS		60249894bp NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4512605	mrna linear EST 14 MAR 2001
DEFINITION		MRNA sequence.	
ACCESSION		BC42B305	
VERSION		BC42B305.1	GI:13334811
KEYWORDS		EST	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Okazaki Y; Motozou S; Chordata; Craniata; Vertebrata; Protostomia;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		1 (bases) 1 to 1037)	
JOURNAL		NIH MGC http://mgs.sci.nih.gov/	
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC).	
		Unpublished (1999)	
		Contact: Robert Traub-Peterson, Ph.D.	
		Email: rpeterson@mail.nih.gov	
		Tissue Procurement: CLONTECH Laboratories, Inc.	
		cDNA library preparation: CLONTECH Laboratories, Inc.	
		cDNA library Arrayed by: The T.M.A.G.E. Consortium (T.M.N.)	
		DNA Sequencing by: Incyte Genomics, Inc.	
		Clone distribution: Mac clone distribution information can be	
		found through the T.M.A.G.E. Consortium/JML at:	
		http://image.jml.gov	
		Plate: LHC1361 row: e column: 22	
		High quality sequence stop: 637.	
FEATURES		Location/Organisms	
Source		1..1037	
		/organism="Homo sapiens"	
		/taxonomy="taxon:9606"	
		/db_xref="IMAGE:4512605"	
		/clone_1fb="NIH_MGC_75"	
		/lab_host="DHIOB (T1 phage-resistant)"	
		/note="Organ: kidney; Vector: pORF-11B (Clontech); Site: 1	
		(gccacccctcaccc); Site: 2: still (ggcattatgac); 5' adapter	
		3' adapters were used in cloning as follows: 5' adaptor	
		sequence: 5'-CATAGCTATTATGGCT-3' and 3' adaptor sequence	
		5'-ATTCTAAGCGCCGACGACGCTGCAATG-dt(40)BR-4' (where B = A,	
		C, or G and N = A, C, G, or T). Average insert size: 1.5 kb	
		kb (range: 0.5-4.0 kb). 15/15 colonies contained inserts	
		by PCR. This library was enriched for full length clones	
		and was constructed by Clontech laboratories (Palo Alto,	
		CA). Note: this is a NIH_MGC library."	
BASE COUNT		292 a	
ORIGIN		218 c	264 t
Query Match	24.0%	Score 497	118 10; Length 1057
Best Local Similarity	100.0%;	Pval. No. 4.9e-246;	
Matches 497; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;





[illegible][illegible]





	KEYWORDS	EST.
	SOURCE	human.
	ORGANISM	Homo sapiens
	REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
	TITLE	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ccgap
	JOURNAL	Unpublished (1997)
	COMMENT	Contact: Robert Strausberg, Ph.D. Email: ccgaps-tr@mail.nih.gov Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Altman, Ph.D., Gerald Marti, M.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <a href="http://www.bio.liv.edu/bicrp/image/image.html">www.bio.liv.edu/bicrp/image/image.html</a> Seq primer: -28ml3 revl ET from Amersham High quality sequence stop: 361
	FEATURES	location/Qualifiers
	SOURCE	1 380
		/organism="Homo sapiens"
		/db_xref="GBR:S945860"
		/db_xref="taxon:9606"
		/clone IMAGE:746336"
		/clone.lib="NCI_CGAP_GCH1"
		/lib.type="germinal center B cell"
		/lab.host="DH10b"
		/note="vector: pTZ19-Pac (pharmacia) with a modified polylinker Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (cd20+, lqo+), provided by Dr. Louis M. Staudt (NCI). Dr. David Altman (NCI) and Dr. Gerald Marti (Ches). cDNA synthesis was primed with a Not I - clipo(dt) primer 15'-GTACCATCATTGAAGTCGGAGCGGCCTATTTTTTTTTTTTTT-3' Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	105 a	90 c 76 g 109 t
ORIGIN		
	Query Match	18.4%; Score 380; DB 9; Length 380.
	Best Local Similarity	100.0%; Pred. No. 1, 9e-185;
	Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DY	854	tatttaaatatgcgtgaagaagcaaacgaactccaactcttgtagacaagtccaggaaagg 913
DB	380	tattgaagaatgccccaagaacaccgaacttcctaactttggtagacacattccacgaagg 321
OY	914	atgatattaatatgaatgaagaatgaagaacctgaacatgaacaaaagcgctactctgaacg 973
DB	320	atttgtttaattgatgaagcatgaagaatccaaacccaaacccaaacccaaacccaaaccc 261
OY	974	gcttatcaccaatgatgatccccgtgaacgaagcgtgaagaactcctgtagctcttggattt 1033
DB	260	gcttgaccaatgatgatgctatcccctgaccccaagcgctggaacattctcattatcttttc 201
OY	1034	gtccacacataaqaatarctatcgtcatatgatgatgtaaagttaatcaaqcaaaarcaaag 1093
DB	200	gtcttcacatcaaacgaactctgctatctaatatggaagctttatcaacgscacaaanamaag 141
OY	1094	tctatacgcgcactataaacactcttgacaacactctctatgacacacacacacacacac 1153
DB	140	tctatatacgcgcactataaacacattcttacacaaactcctctgttaccacacacacacacac 81

OY	1154	cgcagcaccatcttccattacttcctgagtctcaatttgaggatgtttccacacgat	1233
DB	60	ccatgtagcacattttcatgttttacctttctgaatttcacatttggatcttgcacaaataaacat	123
OY	1214	cttgcagaatctctcgaaa	1233
DB	20	tttgcaacatctctggaaaa	1
<hr/>			
RESULT 14			
LOCUS	BE502966	450 bp	mRNA linear EST v4 Aug 2000
DEFINITION	h2b1a07.x1 NCI-CCAP.Ln24 Homo sapiens cDNA clone IMAGE:4214428 similar to FR-075703.075704 HYPOTHETICAL L2.4 KD PROTEIN. mRNA sequence.		
ACCESSION	BE502966		
VERSION	BE502966.1	GI:4705374	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Bukayeva, Metazon; Chordata; Crustacea; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
AUTHORS	1 (bases 1 to 490)		
TITLE	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicdp. National Cancer Institute, Cancer Genome Anatomy Project (CCAP). Tumor Gene Index Unpublished (1997)		
JOURNAL			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rcdbs@r-mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael K. Pomeroy, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL send email too: info@image.lnl.gov Seq primer: 400P from Gibco High quality sequence stop: 427. Location/Qualifiers 1..490 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_1d="IMAGE:3214428" /clone_1id="NCI-CCAP.Ln24" /tissue_type="carcinoid" /lab_host="DH10B" /note="Organ: Lung; Vector: pU714b-Pac (pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CCAP.Ln25 was prepared, and ss cicles were made in vitro. Following BAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneshs 1414420-141991 and 1520904-1522459). Subtraction by Benoit Soares and M. Fatima Bonaldo."		
FEATURES			
SOURCE			
<hr/>			
BASE COUNT	146 a	90 c	101 g 151 t 2 others
ORIGIN			
Query Match	18.1%	Score 375;	ID# 10; Length 450;
Best Local Similarity	100.0%;	Prod. No. 7.8e-185;	
Matches 375;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps 0;			
OY	1155	aattctctctttaaadaaatgttcgatttaadagaaatgtgacgaatcttttgttaaa	214
DB	1	aattctctctttaaadaaatgttcgatttaadagaaatgtgacgaatcttttgttaaa	60
OY	215	tacaaaagatitaaacatatgaatttgatgtttgaagacacacatggaatctgaac	314
DB	61	tcacaaaagatitaaacatatgaatttgatgtttgaagacacacatggaatctgaac	120
OY	315	ataaagatatacagatcacaatcgaatgagttgctagaatttccttctctatctatctat	474









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: July 30, 2002, 03:56:28 ; Search time 58.64 seconds

(without alignments)  
8662.519 Million cell updates/sec

Title: US-09-932-678-1

Perfect score: 1  
Sequence: 1 acagagcgtcgtggtggaag.....tgacatttggaattcccat 2068

Scoring table: GATP 60.0, GATP 60.0

Searched: 383533 seqs, 122816752 residues

Word size: 0

Total number of hits satisfying chosen parameters: 767866

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database:

Issued patents NA:  
1: /cgn2\_5/prodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_5/prodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_5/prodata/1/ina/PCNUS.COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backlist1.seq.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	19	0.9	479 4 US-08-936-165A-192	Sequence 192, App
2	19	0.9	787 1 US-08-034-245-11	Sequence 11, Appl
3	19	0.9	37950 4 US-09-338-907-183	Sequence 183, App
4	19	0.9	37950 4 US-09-218-207-183	Sequence 183, App
5	18	0.9	315 4 US-08-744-419-3	Sequence 3, Appl
6	18	0.9	540 4 US-08-744-419-1	Sequence 1, Appl
7	18	0.9	708 1 US-08-078-090-24	Sequence 24, Appl
8	18	0.9	1065 1 US-08-249-554-1	Sequence 1, Appl
9	18	0.9	1065 1 US-08-249-554-1	Sequence 1, Appl
10	18	0.9	1065 1 US-08-249-554-1	Sequence 1, Appl
11	18	0.9	1065 1 US-08-734-792-1	Sequence 1, Appl
12	18	0.9	1065 1 US-08-078-090-1	Sequence 1, Appl
13	18	0.9	1065 5 PCT-US95-03628-1	Sequence 1, Appl
14	18	0.9	1065 5 PCT-US95-03789-1	Sequence 1, Appl
15	18	0.9	1446 1 US-08-596-024-5	Sequence 5, Appl
16	18	0.9	1446 4 US-09-020-818-5	Sequence 5, Appl
17	18	0.9	1446 4 US-08-907-740-5	Sequence 5, Appl
18	18	0.9	1791 1 US-08-245-294-7	Sequence 7, Appl
19	18	0.9	1791 1 US-08-474-499-7	Sequence 7, Appl
20	18	0.9	1791 5 PCT-US95-06211-7	Sequence 7, Appl
21	18	0.9	6873 4 US-09-131-028A-1	Sequence 1, Appl
22	18	0.9	6873 4 US-09-131-028A-8	Sequence 8, Appl
23	18	0.9	8430 4 US-09-131-028A-6	Sequence 6, Appl
24	18	0.9	8430 4 US-09-131-028A-10	Sequence 10, Appl
25	18	0.9	10607 1 US-08-078-090-3	Sequence 3, Appl
26	18	0.9	10607 1 US-08-078-090-3	Sequence 3, Appl
27	17	0.8	27 3 US-08-468-011A-5	Sequence 5, Appl

28	17	0.8	27 4 US-09-246-468A-5	Sequence 5, Appl
29	17	0.8	27 5 PCT-US95-07085-5	Sequence 5, Appl
30	17	0.8	60 1 US-07-670-246-19	Sequence 19, Appl
31	17	0.8	60 1 US-08-093-781-20	Sequence 20, Appl
32	17	0.8	159 1 US-08-485-4550-78	Sequence 78, Appl
33	17	0.8	159 2 US-08-482-1305-78	Sequence 78, Appl
34	17	0.8	159 2 US-08-484-2115-78	Sequence 78, Appl
35	17	0.8	159 3 US-08-908-769-78	Sequence 78, Appl
36	17	0.8	159 3 US-08-908-769-78	Sequence 78, Appl
37	17	0.8	159 3 US-08-817-795-78	Sequence 78, Appl
38	17	0.8	159 3 US-08-485-4438-78	Sequence 78, Appl
39	17	0.8	159 3 US-08-649-075A-78	Sequence 78, Appl
40	17	0.8	159 4 US-09-012-441-78	Sequence 78, Appl
41	17	0.8	159 4 US-09-012-692-78	Sequence 78, Appl
42	17	0.8	159 4 US-08-906-613-78	Sequence 78, Appl
43	17	0.8	159 5 PCT-US95-14442A-78	Sequence 78, Appl
44	17	0.8	553 4 US-09-227-357-94	Sequence 94, Appl
45	17	0.8	558 2 US-08-896-465-5	Sequence 5, Appl

#### ALIGNMENTS

RESULT 1  
US-08-936-165A-192/c  
Sequence 192, Application US/08936165A  
Patent No. 6348582  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burban, Martin  
APPLICANT: Holston, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nichols, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582e1 Prokaryotic Polynucleotide Uses  
TITLE OF INVENTION: Polypeptides and their uses  
NUMBER OF SEQUENCES: 544  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithkline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24 SEP 1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24 SEP 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R  
REGISTRATION NUMBER: 48,891  
REFERENCE/DOCKET NUMBER: P50549  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-6090  
INDEX:  
INFORMATION FOR SEQ ID NO: 192:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 479 base pairs  
TYPE: nucleic acid  
STRANDNESS: single



GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumentfeld, Maria  
APPLICANT: Ilya, Chumakov  
APPLICANT: Bouaouelert, Lydie  
TITLE OF INVENTION: Prostate cancer gene  
FILE REFERENCE: GENSET.018CP1  
CURRENT APPLICATION NUMBER: US/09/218,207  
CURRENT FILING DATE: 1998-12-22  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
FURTHER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: patent.pm  
SEQ ID NO 183  
LENGTH: 37950  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: exon  
LOCATION: 5259..5328  
OTHER INFORMATION: exon2  
FEATURE:  
NAME/KEY: exon  
LOCATION: 12675..12791  
OTHER INFORMATION: exon3  
FEATURE:  
NAME/KEY: exon  
LOCATION: 14621..14710  
OTHER INFORMATION: exon4  
FEATURE:  
NAME/KEY: exon  
LOCATION: 19822..19912  
OTHER INFORMATION: exon5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 21789..21950  
OTHER INFORMATION: exon6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 23387..23510  
OTHER INFORMATION: exon7  
FEATURE:  
NAME/KEY: exon  
LOCATION: 25520..26016  
OTHER INFORMATION: exon8  
US-09-218-207-183

Query Match 0.98; Score 19; DB 4; Length 37950;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1320 cctctattactataaatt 1338  
Db 1077 cctctattactataaatt 1095

RESULT 5  
US-08-744-419-3  
Sequence 3, Application US/08744419  
Patent No. 6274342  
GENERAL INFORMATION:  
APPLICANT: Gutierrez-Ramos et al.  
TITLE OF INVENTION: Monocyte Chemotactic Protein 5 (MCP-5) Molecules  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA

ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,419  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. Provisional  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: M10-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 315 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-744-419-3

Query Match 0.98; Score 18; DB 4; Length 415;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1824 jctaaagactacaaag 1841  
Db 148 jctaaagactacaaag 155

RESULT 6  
US-08-744-419-1  
Sequence 1, Application US/08744419  
Patent No. 6274342  
GENERAL INFORMATION:  
APPLICANT: Gutierrez-Ramos et al.  
TITLE OF INVENTION: Monocyte Chemotactic Protein 5 (MCP-5) Molecules  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: Boston  
STATE: Massachusetts  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,419  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. Provisional  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: M10-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941



FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
OTHER INFORMATION: The encoded product of nucleotide seq ID NO: 1 is the human  
PUBLICATION INFORMATION:  
AUTHORS: R. Lonnegård et al  
TITLE: Cloning and sequencing of a cDNA encoding human  
JOURNAL: Federation European Biochemical Society Letters  
VOLUME: 269  
ISSUE:  
PAGES: 153 - 156  
DATE: 1990  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-09-249-55A-1

Query Match 0.9%; Score 18; DB 1; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CAGAGCTGTGCTGGA 19  
|||||  
DB 192 CAGAGCTGTGCTGGA 175

## RESULT 9

US-09-249-55A-1/c

Sequence 1, Application US/08249555A

Patent No. 5538952

GENERAL INFORMATION:

APPLICANT: Mukerji, P.

APPLICANT: Seo, A.

APPLICANT: Anderson, S.

APPLICANT: Schaller, J.

TITLE OF INVENTION: Inhibition of Infection of Mammalian Cells by Respiratory Sync

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lonnie R. Drayer

ADDRESSEE: Ross Products Division

ADDRESSEE: Abbott Laboratories

STREET: 625 Cleveland Avenue

CITY: Columbus

STATE: Ohio

COUNTRY: United States

ZIP: 43215

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh System 7.1(D)SOFTWARE: ClarisWorks 1.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/249,555A

FILING DATE: 26-MAY-1994

CLASSIFICATION: 514

PRIOR APPLICATION DATA: NO. 5538952 applicable

TELECOMMUNICATION INFORMATION:

TELEPHONE: (614) 624-3774

TELEFAX: (614) 624-3074

TELEX: No. 5538952e

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1065 base pairs

TYPE: Nucleic acid

STRANDEDNESS: Single

TOPOLOGY: Unknown

MOLECULE TYPE: Cloned cDNA representing the product of a human

MOLECULE TYPE: genomic DNA segment.

DESCRIPTION: Human milk delta-casein

HYPOTHETICAL:

ANTI-SENSE:

FRAGMENT TYPE:

ORIGINAL SOURCE: Human

ORGANISM: Homo sapiens

STRAIN:

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE: Adult

EAPLOTYPE:

TISSUE TYPE: Mammary gland

CELL LINE:

ORGANELLE:

IMMEDIATE SOURCE: Human Mammary Gland

LIBRARY:

CLONE:

POSITION IN GENOME:

CHROMOSOME/SEGMENT:

MAP POSITION:

UNITS:

FEATURE:

NAME/KEY:

LOCATION:

IDENTIFICATION METHOD: DNA sequencing and restriction analysis

OTHER INFORMATION: The encoded product of nucleotide seq ID NO: 1 is the human

PUBLICATION INFORMATION:

AUTHORS: R. Lonnegård et al

TITLE: Cloning and sequencing of a cDNA encoding human

TITLE: milk beta-casein.

JOURNAL: Federation European Biochemical Society Letters

VOLUME: 269

ISSUE:

PAGES: 153 - 156

DATE: 1990

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

US-09-249-55A-1

Query Match 0.9%; Score 18; DB 1; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CAGAGCTGTGCTGGA 19  
|||||  
DB 192 CAGAGCTGTGCTGGA 175

## RESULT 10

US-09-249-584-1/c

Sequence 1, Application US/08249584

Patent No. 5644880

GENERAL INFORMATION:

APPLICANT: Mukerji, P.

APPLICANT: Seo, A.

APPLICANT: Anderson, S.

TITLE OF INVENTION: Product for Inhibition of Attachment of

TITLE OF INVENTION: H. Influenzae to Human Cells.

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lonnie R. Drayer

ADDRESSEE: Ross Products Division

ADDRESSEE: Abbott Laboratories

STREET: 625 Cleveland Avenue

CITY: Columbus

STATE: Ohio

COUNTRY: United States

ZIP: 43215

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb storage

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh System 7.1

SOFTWARE: CLUSTALWSK 1.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/249,584  
 FILING DATE: 26-MAY-1994  
 CLASSIFICATION: 4.24  
 PUBLICATION INFORMATION DATA: No. 964880 applicable  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (614) 624-3774  
 TELEFAX: (614) 624-4074  
 INDEX: No. 664880  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1065 base pairs  
 TYPE: Nucleic acid  
 STRANDEDNESS: Single  
 TOPOLOGY: Unknown  
 MODIFIED TYPE: cloned cDNA representing the product of a human  
 MODIFIED TYPE: genomic DNA segment  
 DESCRIPTION: Human milk 'cld' casin  
 ORIGIN: HUMAN  
 ANTI-SENSE:  
 PRESENT TYPE:  
 ORGANISM: Homo sapiens  
 STRAIN:  
 INDIVIDUAL ISOLATE:  
 DEVELOPMENTAL STAGE: Adult  
 TISSUE TYPE: Mammary gland  
 CELL TYPE:  
 CELL LINE:  
 ORGANELLER:  
 IMMEDIATE SOURCE: Human Mammary gland  
 TISSUE:  
 POSITION IN GENOME:  
 ORIGIN/SEQUENCE/SEGMENT:  
 MAP POSITION:  
 UNIT:  
 FEATURE:  
 NAME/KEY:  
 IDENTIFICATION METHOD: DNA sequencing and restriction  
 IDENTIFICATION METHOD: analysis  
 OTHER INFORMATION: The cloned product of nucleotide  
 OTHER INFORMATION: SEQ ID NO:1 is the human milk protein,  
 OTHER INFORMATION: "casin".  
 INFORMATION INFORMATION:  
 AUTHORS: R. Lomax et al.  
 TITLE: Cloning and sequencing of a cDNA encoding human  
 TITLE: milk beta-casin  
 JOURNAL: European Biochemical Society Letters  
 VOLUME: 269  
 PAGES: 153-156  
 DATE: 1990  
 JOURNAL NUMBER:  
 FILING DATE:  
 PUBLICATION DATE:  
 RELEVANT RESIDUES IN SEQ ID NO:1  
 OR: 249,584-1

US 08 734-792-1/C  
 Sequence 1, Application US/08/249,584  
 Patent No. 5707668  
 GENERAL INFORMATION:  
 APPLICANT: Mokeji, P.  
 APPLICANT: Seo, A.  
 APPLICANT: Anderson, S.  
 APPLICANT: Harvey, L.  
 TITLE OF INVENTION: Inhibition of Attachment of H. Influenzae  
 TITLE OF INVENTION: to Human Cells  
 NUMBER OF SEQUENCES: 5  
 CROSS-REFERENCE ADDRESS:  
 ADDRESSER: Lante R. Drayet  
 ADDRESSER: Ross Products Division  
 ADDRESSER: Abbott Laboratories  
 STREET: 625 Cleveland Avenue  
 CITY: Columbus  
 STATE: Ohio  
 COUNTRY: United States  
 ZIP: 43215  
 COMPUTER READABLE FORM:  
 SEQUENCE TYPE: 4.5 inch, 1.44 Mb storage  
 COMMENTER: Apple Macintosh  
 SEPARATING SYSTEM: Macintosh system 7.1  
 SOFTWARE: CLUSTALWSK 1.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/734,792  
 FILING DATE: 24-MAY-1994  
 CLASSIFICATION: 5.14  
 PUBLICATION INFORMATION DATA:  
 APPLICATION NUMBER: 08/249,584  
 FILING DATE: 26-MAY-1994  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (614) 624-3774  
 TELEFAX: (614) 624-4074  
 INDEX: No. 5707668  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1065 base pairs  
 TYPE: Nucleic acid  
 STRANDEDNESS: Single  
 TOPOLOGY: Unknown  
 MODIFIED TYPE: cloned cDNA representing the product of a human  
 MODIFIED TYPE: genomic DNA segment  
 DESCRIPTION: Human milk 'cld' casin  
 ORIGIN: HUMAN  
 ANTI-SENSE:  
 PRESENT TYPE:  
 ORGANISM: Homo sapiens  
 STRAIN:  
 INDIVIDUAL ISOLATE:  
 DEVELOPMENTAL STAGE: Adult  
 TISSUE TYPE: Mammary gland  
 CELL TYPE:  
 CELL LINE:  
 ORGANELLER:  
 IMMEDIATE SOURCE: Human Mammary gland  
 TISSUE:  
 POSITION IN GENOME:  
 ORIGIN/SEQUENCE/SEGMENT:  
 MAP POSITION:  
 UNIT:  
 FEATURE:  
 NAME/KEY:  
 IDENTIFICATION METHOD: DNA sequencing and restriction  
 IDENTIFICATION METHOD: analysis  
 OTHER INFORMATION: The cloned product of nucleotide  
 OTHER INFORMATION: SEQ ID NO:1 is the human milk protein, "casin".  
 INFORMATION INFORMATION:

RESULT 11



AUTHORS: B. Lommerdal et al  
TITLE: Cloning and sequencing of a cDNA encoding human  
TITLE: milk beta-casein  
JOURNAL: Federation European Biochemical Society Letters  
VOLUME: 269  
ISSUE:  
PAGES: 153 - 156  
DATE: 1990  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-734-792-1

Query Match 0.9%; Score 18; DB 1; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 caaagctgtgctgcaaa 19  
|||||  
DB 192 CAGAGCTGTGCTGCAAA 175

RESULT 12  
US-08-078-090-1/c  
Sequence 1, Application US/08078090  
Patent No. 573407  
GENERAL INFORMATION:  
APPLICANT: BERGSTROM, SVEN  
APPLICANT: HERNELL, OLE  
APPLICANT: LOENNERDAL, BO  
APPLICANT: HJALMARSSON, KARIN  
APPLICANT: HANSSON, LENNART  
APPLICANT: TOERNELL, JAN  
APPLICANT: STROMQVIST, MATS  
TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING  
NUMBER OF INVENTIONS: 17 AND USE THEREOF  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSER: BROKEDY AND NEIMARK  
STREET: 419 SEVENTH STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/078-090  
FILING DATE: 19930618  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/DK92/00236  
FILING DATE: 19-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/DK91/00233  
FILING DATE: 19-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P.  
REGISTRATION NUMBER: 28 005  
REFERENCE/DOCKET INFORMATION:  
TELEPHONE: (202)628-5197  
TELEFAX: (202)737-3528  
TELEFAX: 248633  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: s14-peptide  
LOCATION: 4.48  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 49..681  
US-08-078-090-1

Query Match 0.9%; Score 18; DB 1; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 caaagctgtgctgcaaa 19  
|||||  
DB 192 CAGAGCTGTGCTGCAAA 175

RESULT 13  
PCT-US-95-03628-1/c  
Sequence 1, Application PCT/US9503628  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories  
TITLE OF INVENTION: Inhibition of Infection of Mammalian Cells by  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSER: LORLIE R. DRYER  
ADDRESSER: ROSS Products Division  
ADDRESSER: Abbott Laboratories  
STREET: 625 Cleveland Avenue  
CITY: Columbus  
STATE: Ohio  
COUNTRY: United States of America  
ZIP: 43215  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 4.5 inch, 1.44 Mb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS Version 6.21  
SOFTWARE: WordPerfect Version 6.0a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/03628  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/249,554  
FILING DATE: 26-MAY 1994  
APPLICATION NUMBER: US 08/249,555  
FILING DATE: 26-MAY-1994  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-3774  
TELEFAX: (614) 624-3074  
TELEX: None  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: single  
MOLECULE TYPE: Cloned cDNA representing the product of a  
MOLECULE TYPE: human genomic DNA sequent  
DESCRIPTION: Human milk eta-casein  
HYPOTHEICAL:  
ANTI-SENSE:  
FEATURE:  
NAME/KEY: s14-peptide  
ORIGINAL SOURCE: human  
ORGANISM: Homo sapiens  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE: Adult

BALTIC TYPE: Mammmary gland  
 CELL TYPE:  
 CELL LINE:  
 ORGANISM:  
 IMMEDIATE SOURCE: Human Mammmary gland  
 LIBRARY:  
 CLONE:  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT:  
 MAP POSITION:  
 UNITS:  
 FEATURE:  
 NAME/KEY:  
 LOCATION:  
 IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
 OTHER INFORMATION: The encoded product of nucleotide SEQ ID  
 OTHER INFORMATION: No. 1 is the human milk protein, *casin*.  
 PUBLICATION INFORMATION:  
 AUTHOR: H. Lammert, et al.  
 TITLE: Cloning and sequencing of a cDNA encoding human milk beta-  
 CASEIN.  
 JOURNAL: Federation European Biochemical Society Letters  
 VOLUME: 269  
 ISSUE:  
 PAGES: 153 - 156  
 DATE: 1990  
 DOCUMENT NUMBER:  
 FILING DATE:  
 PUBLICATION DATE:  
 RELEVANT RESIDUES IN SEQ ID NO:

Query Match 0.98; Score 18; 108 5; Length 1065;  
 best local similarity 100.0%; Prod. No. 24;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 cagagatgagctgaa 19  
 DB 192 cagagctgctgctgaa 179

RESULT 14  
 Seq-ID# 03789 1/0  
 Seq-ID# 1; Application no/US9504789  
 GENERAL INFORMATION:  
 APPLICANT: Abbott Laboratories  
 TITLE OF INVENTION: Inhibition of Attachment of H. Influenzae to Human  
 TITLE OF INVENTION: Cells  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENT ADDRESS:  
 ADDRESSEE: Jounie R. Drayer  
 ADDRESSEE: Ross Products Division  
 ADDRESSEE: Abbott Laboratories  
 STREET: 625 Cleveland Avenue  
 CITY: Columbus  
 STATE: Ohio  
 COUNTRY: United States of America  
 ZIP: 43215  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 4.5 inch, 1.44 MB storage  
 COMMENTS: IBM compatible  
 OPERATING SYSTEM: MS DOS Version 6.21  
 SOFTWARE: WordPerfect Version 6.0a  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/04789  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/249,556  
 FILING DATE: 26 MAY 1994  
 APPLICATION NUMBER: US 08/249,584

FILING DATE: 26 MAY 1994  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (614) 624-4774  
 TELEFAX: (614) 624-4074  
 TELEX: None  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1065 base pairs  
 TYPE: Nucleic acid  
 STRANDEDNESS: Single  
 TOPOLOGY: Unknown  
 MODIFIED TYPE: Cloned cDNA representing the product of a  
 MULTIPLE TYPE: human genomic DNA segment  
 DESCRIPTION: Human milk *eta* casein  
 HYPOTHETICAL:  
 ANTI-SENSE:  
 FRAGMENT TYPE:  
 ORIGINAL SOURCE: Human  
 ORGANISM: Homo Sapiens  
 STRAIN:  
 INDIVIDUAL ISOLATE:  
 DEVELOPMENTAL STAGE: Adult  
 BALLOTYPE:  
 TISSUE TYPE: Mammmary gland  
 CELL TYPE:  
 CELL LINE:  
 ORGANISM:  
 IMMEDIATE SOURCE: Human Mammmary gland  
 LIBRARY:  
 CLONE:  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT:  
 MAP POSITION:  
 MAPS:  
 FEATURE:  
 NAME/KEY:  
 LOCATION:  
 IDENTIFICATION METHOD: DNA sequencing and restriction analysis  
 OTHER INFORMATION: The encoded product of nucleotide SEQ ID  
 OTHER INFORMATION: No. 1 is the human milk protein, *casin*.  
 PUBLICATION INFORMATION:  
 AUTHOR: H. Lammert, et al.  
 TITLE: Cloning and sequencing of a cDNA encoding human milk beta-  
 CASEIN.  
 JOURNAL: Federation European Biochemical Society Letters  
 VOLUME: 269  
 ISSUE:  
 PAGES: 153 - 156  
 DATE: 1990  
 DOCUMENT NUMBER:  
 FILING DATE:  
 PUBLICATION DATE:  
 RELEVANT RESIDUES IN SEQ ID NO:

Query Match 0.98; Score 18; 108 5; Length 1065;  
 best local similarity 100.0%; Prod. No. 24;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 cagagatgagctgaa 19  
 DB 192 cagagctgctgctgaa 179

RESULT 15  
 Seq-ID# 03789 1/0  
 Seq-ID# 5; Application no/US0896024  
 GENERAL INFORMATION:  
 APPLICANT: Barry Gerard P.  
 APPLICANT: Demoyd, Jan W.  
 APPLICANT: Kishore, Ganesh M

APPLICANT: Weidon, Marcia L.  
TITLE OF INVENTION: Expression of Sucrose Phosphorylase in  
TITLE OF INVENTION: Plants  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Charles E. Cohen, Monsanto Company, B44F  
STREET: 700 Chesterfield Parkway No. 5716837th  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM pc compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,024  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/386,860  
FILING DATE: 10-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: COHEN, Charles E.  
REGISTRATION NUMBER: 34,565  
REFERENCE/DOCKET NUMBER: 38-21(1367)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)537-6224  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ. ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1446 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-596-024-5

Query Match 0.98; Score 18; PH 1; Length 1446;  
Fast Local Similarity 100.0%; Pred. No. 24;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 459 gctttcttqtaactt 476  
|||||  
DB 304 GCTTTCTGCTAATCTT 287

Search completed: July 30, 2002, 05:14:10  
Job time: 4662 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 03:59:49 ; Search time 243.92 Seconds

(without alignments)  
14556.326 Million cell updates/sec

Title: US-09-932-678-1  
Perfect score: 2068  
Sequence: 1 acagagcgtctgtgctggaag.....tgcacattggtattcccat 2068

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq.012802:\*

```

1: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL/gcgcdata/geneseq/geneseq-emb1/NA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1869	90.4	2040	21	Human RNA polymera
2	920	44.5	1418	21	Human ORF2492
3	842	40.7	1498	22	Human full-length
4	691	33.4	1423	22	Human secretory mo
5	292	14.1	341	20	EST clone D1149
6	269	13.0	3169	23	DNA encoding novel
7	232	11.2	1461	22	Human transcriptio
8	230	11.1	2103	21	Human transcrip
9	229	11.1	2991	23	DNA encoding novel

10	229	11.1	2991	24	AA591992	DNA encoding novel
11	223	10.8	776	24	AA592253	DNA encoding novel
12	215	10.4	3175	24	AA592254	DNA encoding novel
13	208	10.1	34269	22	AAK68677	Human immune/bacma
14	208	10.1	34269	22	AAK85168	Human immune/bacma
15	198	9.6	2410	24	AA592255	DNA encoding novel
16	196	9.5	437	20	AA51664	Human secretory pro
17	190	9.2	396	23	AA580954	DNA encoding novel
18	190	9.2	396	23	AA587635	DNA encoding novel
19	190	9.2	396	23	AA591986	DNA encoding novel
20	190	9.2	396	23	AA592246	DNA encoding novel
21	187	9.0	17803	22	AAK68676	Human immune/bacma
22	182	8.8	498	23	AA592247	DNA encoding novel
23	180	8.7	349	21	AA577533	Human ORF2492
24	176	8.5	9259	22	AAK68674	Human immune/bacma
25	176	8.5	23815	22	AAK68678	Human immune/bacma
26	176	8.5	24815	22	AAK85169	DNA encoding novel
27	175	8.5	585	24	AA592248	DNA encoding novel
28	174	8.4	1017	24	AA587646	DNA encoding novel
29	139	6.7	655	23	AA580955	DNA encoding novel
30	139	6.7	1260	23	AA587643	DNA encoding novel
31	133	6.4	627	23	AA587640	DNA encoding novel
32	117	5.7	213	21	AA519091	Human secretory pro
33	110	5.3	3410	22	AAK85167	Human immune/bacma
34	99	4.8	560	24	AA591991	DNA encoding novel
35	91	4.4	1840	22	AAK94394	Human full-length
36	80	3.9	405	22	AAK56711	Human immune/bacma
37	76	3.7	414	22	AAK6215	Human breast cell
38	76	3.7	414	22	AAK56760	Human breast cell
39	76	3.7	414	22	AAK26382	Human breast cell
40	76	3.7	414	22	AAK04888	Human brain expres
41	76	3.7	414	22	AAK30411	Human bone marrow
42	76	3.7	414	22	AA115019	Protein 4452 for q
43	76	3.7	414	22	AA136361	Protein 5047 used f
44	76	3.7	414	22	AA104785	Protein 4476 used f
45	75	3.6	95	22	ABAS1325	Human breast cell

#### ALIGNMENTS

```

RESULT 1
AAA98384 standard: cDNA, 2040 bp.
ID AAA98384:
XX
AC AAA98384:
XX
BT 08 FEB-2001 (first entry)
XX
DE Human RNA polymerase 1 transcription factor TIF-1A cDNA.
XX
KW RNA polymerase 1 transcription factor TIF-1A; antitumor; treatment;
antiproliferative; cell proliferation; cancer; tissue regeneration; SS.
XX
OS Hc-29 sapiens.
XX
PN W020055316-A1.
XX
PD 21 SEP-2000.
XX
PE 08 MAR-2000: 2000W0-TR00767.
XX
PF 17 MAR-1999: 99DE-1011992.
XX
PA (DEKR-) DEUT KRENSFORCHUNGSZENTRUM.
XX
PI Grunert I, Vinograd M:
XX
WP1: 2000-587527/55.
XX
DR P-I-SDB: AAB10936.
XX
PT Nk* RNA encoding the transcription factor TIF-1A, useful for preventing
or treating diseases associated with abnormal cell proliferation.

```

$$\begin{aligned} \text{For } \mathbf{r} &= (\mathbf{r}^0, \mathbf{r}^1, \mathbf{r}^2, \mathbf{r}^3) \in \mathbb{R}^4, \quad \mathbf{r}^0 = \mathbf{0} \\ \mathbf{r}^1 &= \mathbf{r}^0 + \mathbf{r}^1 \mathbf{e}_1, \quad \mathbf{r}^2 = \mathbf{r}^1 + \mathbf{r}^2 \mathbf{e}_2, \quad \mathbf{r}^3 = \mathbf{r}^2 + \mathbf{r}^3 \mathbf{e}_3 \end{aligned}$$

# THE UNIVERSITY OF CHICAGO

These hybridization results are consistent with the DNA sequence (1) that encodes the RNA

polymers of the transfer factor TTF-1A which has an antitumor activity and collagenase-inducing activity. The

host fibers (I) RNA (Ia) enclosing a protein (II) with the biological

able to bind specifically to, and cleave, its transcribed RNA so

product of hnRNP synthesis of the corresponding protein, (3) an effect of hnRNP RNA (AS) with binding properties similar to R; (4) an

expression vector that contains (1), (1a) or sequences that encode R<sub>1</sub> and R<sub>2</sub> (2), (2a) and (2b). The  $\Delta$ OR (11) and  $\Delta$ OR (15) host cells contain the vectors of (4a) and (6). The  $\Delta$ OR (11)

provided by (1) or (1a); (7) preparation of THF-1A or (11) by culturing

that weakened or blocked the activity of PIF-1A or (11); (10) a diagnostic

of  $\gamma$  and the method in (10), (11) and similar sequences that ensure











CC	disordersinvolving aberrant protein expression or biological activity
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at <a href="http://wipo.int/pub/published_pcl_sequences">http://wipo.int/pub/published_pcl_sequences</a> .
XX	
SO	Sequence 3169 BP; 812 A; 841 C; 779 G; 737 T; 0 other:
XX	
Query Match	13.0%; Score 269; DB 23; Length 3169;
Best local Similarity	100.0%; Pred. No. 4,4e-119;
Matches 269; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1245 ccaagaatcctccatccatccagcagcctctgaataatataatgaacgttttggca 1304
DB	878 ccaadgaatccctccatccatccagcagcctctgaataatataatgaacgttttggca 937
QY	1305 aqagctaaattatctctcttatactgtgaataatcatcctaagctctttgattaaatga 1364
DB	938 aqagctaaattatctctcttatactgtgaataatcatcctaagctctttgattaaatga 997
QY	1365 ctgacataacttaataaacagagattccggaacaagcattctgacatgtgtcttc 1424
DB	998 ctgacataacttaataaacagagattccggaacaagcattctgacatgtgtcttc 1057
QY	1425 catggacatttaactcaactctgcaagctgtgtcttaacacgttttllagaacaaag 1484
DB	1058 catggacatttaactcaactctgcaagctgtgtcttaacacgttttllagaacaaag 1117
QY	1485 caagcttttggcggaacactgaagaagaag 1513
DB	1118 caagcttttggcggaacactgaagaagaag 1146
RESULT 7	
ABAB3038	
ID	ABAB3038 standard; DNA; 1461 BP.
XX	
AC	ABAB3038;
XX	
DT	05-FEB-2002 (first entry)
XX	
DE	Human transcription factor TRFX-65 coding sequence.
XX	
KW	Human: transcription factor; TRFX: cell proliferative disease;
KW	autoimmune disease; inflammation; neurological disease;
KW	developmental disorder; cancer; AIDS: infection; cytostatic; anti-HIV;
KW	neuroprotective; antiinflammatory; gene therapy; ds.
XX	
XX	Homo sapiens.
XX	
XX	W0200172777-A2.
XX	
XX	04-CT-2001.
XX	
XX	13-MAR-2001; 2001WO-US08117.
XX	
XX	13-MAR-2000; 2000US-0188986.
XX	
XX	(INCY-) INCYTE GENOMICS INC.
XX	
XX	Hillman JL, Baughn MR, Yue H, Lal P, Lu DM, Patterson G;
XX	Pi Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;
XX	Reddy K;
XX	WPI: 2001-570896/64.
XX	
XX	P-PSDB, ABB50214.
XX	

[illegible]





















Db 1681 CAGTATTAGAGTACCGCTGAGAGACTCACTGACATCTGCACAAACCCCTTCACACA 1740  
 QY 1741 cctcttcccttctgactcctgtgctgagagagtcgaagaatctatgactctaltt 1800  
 Db 1741 cctcttcccttctgactcctgtgctgagagagtcgaagaatctatgactctaltt 1800  
 QY 1801 atcagatag 1860  
 Db 1801 ATCAGATAG 1860  
 QY 1861 aagacatag 1920  
 Db 1861 AGACATAG 1920  
 QY 1921 cctgattatgagatgag 1980  
 Db 1921 cctgattatgagatgag 1980  
 QY 1981 tggagtcctccacccctgttctacatgagagagagagagagagagagagagagag 2040  
 Db 1981 TGGGTCCTCCACCCGCTGTGTACATGCAACCCGCTCCCTGAGCGGAGAAATTTCGA 2040  
 QY 2041 ctgagatgag 2068  
 Db 2041 CTGAGATGAG 2068

RESULT 2  
 HSA272050 3756 bp mRNA linear PRI 07-SEP-2000  
 LOCUS HSA272050  
 DEFINITION Homo sapiens mRNA for transcription initiation factor 1A protein  
 (TIF-1A gene).  
 ACCESSION AJ272050.1 GI:10046713  
 VERSION AJ272050.1  
 KEYWORDS TIF-1A; transcription initiation factor.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homini; Homo.  
 REFERENCE 1 (bases 1 to 3756)  
 Bodem, J., Hoffmann-Kohrer, U., Ross, W., Delius, H., Vianzon, M. and  
 Grunmt, I.  
 Cloning and functional characterization of transcription initiation  
 factor TIF-1A, a growth-dependent regulator of ribosomal RNA  
 synthesis  
 unpublished  
 2 (bases 1 to 3756)  
 Bodem, J.J.  
 Direct Submission  
 Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle  
 11, Deutsches Krebsforschungszentrum, INF. 280, 69120 Heidelberg,  
 GERMANY

FEATURES  
 source location/Qualifiers  
 1..3756  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 23..3756  
 /gene="TIF-1A"  
 23..1978  
 /gene="TIF-1A"  
 /function="transcription initiation factor"  
 /codon\_start=1  
 /evidence="experimental"  
 /product="transcription initiation factor 1A protein"  
 /protein\_id="CA07955.1"  
 /db\_xref="GI:10046714"  
 /translation="MAAPILHRLPGDAASSAVKKIIGASRTGISMKRALENDPENS  
 PPRIVFGSTVEVLEILKYKGETNDFELKNQIDPDKDOGLINLSEHPSPV  
 TKPEHOLISILRLPMVNSQTEVEYEAFLCNLSADIVLRLGLSMLASHPEPV  
 LKGGVDVSDSDNDNLPAINDCHRALDIIVAVYSTPFWLMPILVKEPEVRS  
 ERTLEGVNHLRLISVPEPTLRLHELELELELELELELELELELELELELELELELE  
 TSTTEGLFNNDEDELEHELETKAPERLDMPPAERLDITKSTVISTYMDCCVTVCK  
 VDNCKTKULYRDLNIFKLLPLTHASCHVCPMYLCSFRLGFAENFLHLMKKIDQ

Query Match 92.8%; Score 1920; DB 9; Length 3756;  
 Best local similarity 99.9%; Pred. No. 0;  
 Matrices 2020; Conservative 0; Mismatches 2; Indels 0; Gaps 0

3756R /gene="TIF-1A"  
 EVIDENCE="experimental"  
 BASE COUNT 1060 a 692 c 844 g 1156 t 4 others  
 ORIGIN

1 NSNPAIIPGAGNVTGSLIAARFIPLIIPKSLDILVIMLILYINQNSIKRATV  
 ALHGPPYSAQVAFYFVFRKQILSGNREKGLVQIOLINFERIVSOLNPLKILRS  
 VVFFPAIITNKOLVECTYIIIRNNRQMLPVIHSTAGNSVGLTNPIDTPPRPVV  
 LKRSKEIDIPYVWEDKSAEELDEFKPKMKDIVEREDIDFLKGVPRNDIVLGLTP  
 SEDTHRENSSSVSGPVLVMSPL\*  
 1979 - 3756

47 tggatcctgagatgag 106  
 Db 1 cctgctcctgagatgag 60  
 QY 107 tggagtcctccacccctgttctacatgagagagagagagagagagagagagagag 166  
 Db 61 tggagtcctccacccctgttctacatgagagagagagagagagagagagagagag 120  
 QY 167 tatgagatgag 226  
 Db 121 tatgagatgag 180  
 QY 227 tggagtcctccacccctgttctacatgagagagagagagagagagagagagagag 286  
 Db 161 tggagtcctccacccctgttctacatgagagagagagagagagagagagagagag 240  
 QY 287 gttgag 346  
 Db 241 gttgag 300  
 QY 347 ggaatcctgttctacatgagagagagagagagagagagagagagagagagagagag 406  
 Db 301 ggaatcctgttctacatgagagagagagagagagagagagagagagagagagagag 360  
 QY 407 attgagatgag 466  
 Db 361 attgagatgag 420  
 QY 467 ggaatcctgttctacatgagagagagagagagagagagagagagagagagagagag 526  
 Db 421 ggaatcctgttctacatgagagagagagagagagagagagagagagagagagagag 480  
 QY 527 cctatgag 586  
 Db 481 cctatgag 540  
 QY 587 tggagtcctccacccctgttctacatgagagagagagagagagagagagagagag 646  
 Db 541 tggagtcctccacccctgttctacatgagagagagagagagagagagagagagag 600  
 QY 647 aagacatag 706  
 Db 601 aagacatag 660  
 QY 707 atttcttgcgaatgag 766  
 Db 661 atttcttgcgaatgag 720  
 QY 767 tggatcctccacccctgttctacatgagagagagagagagagagagagagagag 826  
 Db 721 tggatcctccacccctgttctacatgagagagagagagagagagagagagagag 780  
 QY 827 gttgagatgag 886  
 Db 781 gttgagatgag 840  
 QY 887 ttgtgag 946



QY	1447	aaagatcccttctctctatccatgcgaacttgcgaagaaagacttgaagcaacttatacgaattat	406
DB	318	AGAAATTCGGTCTCTCTCTATACATGACTTCGAAATAAAGACCTTTGAGCAATATACGATATAT	377
QY	407	attaagattgccttgatttgaataaaagctaaacagtaagtgaagaaagatttgaacttctct	466
DB	378	ATTAAATATTCGCTTGTTGGATAGAAAGTAAACAGTACTCGAACAGATATTTGGCTTTCT	437
QY	467	tgatataccttgatcagacaaagacttttctcagacagctctcagactgaattgcttc	526
DB	438	TGGTATATCTTGATATCAGACAGACTGTTTTCTCGACACCGTGTCTGACGACATTCCTTC	497
QY	527	ccatttttgccttcccagatgatcattaagaagcgaatgtagatgttcagattctcga	586
DB	498	CAATTTTGGCTCCCGACATGACATATAGGAAGCGCATATACATGTTTCAATTCTCGA	557
QY	587	tgatgaagatgatactctctcgaattttgaacatgctacagagcttgcgaataat	646
DB	558	TGATGAGATATGATATCTTCTCTGCAAAATTTGACACATGTCACAGACCTTGGAATAT	617
QY	647	aagaaagatgcacatctcgaagacccttgattctctatgcgaatacttgatgaagaaattcc	706
DB	618	AGCAAAATATATGACCATTCGAAACCGGTGGTTTCTCATGTCGAATACGTGTGAAAAATTCG	677
QY	707	atttgctcgaataatcgaagaaagactcgaactgatacttataactactaagattag	766
DB	678	ATTGTGTGCAAAATCAGACAGAACACGTGATCTTACCTTCATACCTACTTAAAGATTAG	737
QY	767	tgatatttcccaacttgaagcctgaatacttcgaagctatcttctgaaacttacttca	826
DB	738	TGATATTTTTCACATCTTGAGCATTAATAATCTTGAGGCTTATATTTCAAAAAATACTGAA	797
QY	827	gttgagatggaatgcattccgcgaaggtatlaaagatgctgaagaaacagcaactcaac	886
DB	798	GTTGCAATGGAATGCATTCGCGCAGGGATATGAACATCTTCMAAACAACATCTCAAC	857
QY	887	tttgatgtaggaanaattcgcgaagaaagattgtttaalaagatgaagaaatgaagaaacga	946
DB	858	TTTGCTGTGGACACATTCACAGGAAGATTTGTTTAAATATGATGATGAAGTAAAGAAATCTGA	917
QY	947	acatgaacaaagcctggtcctgaacagcctcgaacagaatgctgcatcctgaacccagcg	1006
DB	918	ACATGGAACAAAGCGCTGATCTCTTAAACAGTCCAGCAAGATGCTGATCTTACCCAGCG	977
QY	1007	cctgagacatccctgagtgctttgttttgcctcaagtaagaaatgctcgaatgtagatg	1066
DB	978	CTTGACATCTCGATGTTTTTGCTTTGTCTTACATGAAGAGTGTCTGCTATGTACATG	1037
QY	1067	taagatgataacgccaanaaagaatctatatgcgaactgaataaactcttgaata	1126
DB	1038	TAAGCTTCAATACGCAAAAGCAAAAGCATTTAATTCGACGCTGATMAACATCTTCACAA	1097
QY	1127	actcctgtgcccacccctgcctccctgcgcacatgaagtttttgaagtgttaactcctgag	1186
DB	1098	ACTTCGTGTGGCCACCCATGCTCTCTGCGATATACAGTTTTCATGTTTTTACCTCTGAG	1157
QY	1187	tctcaaatcgaatctgcagagcaatctttgaaacatctcgaagaaatctgcagacc	1246
DB	1158	TTTCAAAATTCGATTCGACAGAGCATTTTTCGACACATCTCTGCAAAAAATATCGAGACCC	1217
QY	1247	aagatacctgcacatcgaagcagcgtctcgaattataltatgaagccttttcgaag	1306
DB	1218	AATATATCTGTCATATCAGACAGCGCTGCGAAAAATTATTTTGCAGACCTTTTGGCAAG	1277
QY	1307	agcttaattalacctctatacgaataaactcactgcctagatcctttgattaaactgct	1366
DB	1278	AGCTAAATTTATCTCTTATATGCTTAAATATCACTGCTACATCTTTTGGTTAACTGACT	1337
QY	1367	gacacatctccttaataaccaggaatctcgaagaaaggaatctcgaacatttactctca	1426
DB	1338	GCACATATACCTTATATACGAGCATTCGGGAAACAAGGCATTTCGCACTTTCTCTGCA	1397
QY	1427	tgagacatttactcagcctcgaagcgtcttctatacccttgatttgaacacaga	1486

[illegible]









\* arbitrary gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1033: contig of 1013 bp in length  
 \* 1034 1133: gap of unknown length  
 \* 1134 2547: contig of 1414 bp in length  
 \* 2548 2647: gap of unknown length  
 \* 2648 4205: contig of 1558 bp in length  
 \* 4206 4305: gap of unknown length  
 \* 4306 6564: contig of 2259 bp in length  
 \* 6565 6664: gap of unknown length  
 \* 6665 9065: contig of 2401 bp in length  
 \* 9066 9165: gap of unknown length  
 \* 9166 12059: contig of 2894 bp in length  
 \* 12060 12159: gap of unknown length  
 \* 12160 19295: contig of 7136 bp in length  
 \* 19296 19395: gap of unknown length  
 \* 19396 25668: contig of 6273 bp in length  
 \* 25669 31013: gap of unknown length  
 \* 31014 31113: contig of 5245 bp in length  
 \* 31114 36555: contig of 5442 bp in length  
 \* 36556 36655: gap of unknown length  
 \* 36656 45388: contig of 8683 bp in length  
 \* 45389 45439: gap of unknown length  
 \* 45440 52918: contig of 7480 bp in length  
 \* 52919 53018: gap of unknown length  
 \* 53019 65720: contig of 12702 bp in length  
 \* 65721 65821: gap of unknown length  
 \* 65822 76109: contig of 10289 bp in length  
 \* 76110 76209: gap of unknown length  
 \* 76210 85439: contig of 9230 bp in length  
 \* 85440 101620: gap of unknown length  
 \* 101621 101720: gap of unknown length  
 \* 101721 122163: contig of 20443 bp in length  
 \* 122164 122264: gap of unknown length  
 \* 122265 143900: contig of 21637 bp in length.

FEATURES  
 source  
 1. 143900  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-344H15"  
 /clone\_lib="RP11 human BAC library 11"

BASE COUNT 17791 a 34128 c 34320 g 35886 t 1775 others  
 ORIGIN

Query Match 10.1%; Score 208; DB 2; Length 143900;  
 Best Local Similarity 100.0%; Pred. No. 4e-98;  
 Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1861 aagagatcagagagatgagatgactttctgaaagcgaatgcacacaaataata 1920  
 |||||||  
 Db 27338 AGGACATGAGTGGAGATGAGATGACTTTCTGAAGAGCCAGATGCGCCAGATGTA 27279  
 QY 1921 cctgattgagatcacacgaagctccttgcacgcatcttcgaagtccttcaagtaag 1980  
 |||||||  
 Db 27278 CGGTGATTGGGATCAGACGCAAGCTCCTTGACAGCATTTCCGCAATGCTTCAATATAC 27219  
 QY 1981 tgggttcctccacccggttggtaacatgcaacacgctcccttcgacgcaagaattatga 2040  
 |||||||  
 Db 27218 TGGGCTCCCGACCCGCTGTGTGACATGCAACCCAGTCCCTCGACGCGAGAAATTTCGA 27159  
 QY 2041 ctgagatgacatcttgagatcccat 2068  
 |||||||  
 Db 27158 CTGACATGTGACATTTGGCATTCGCCAT 27131

RESULT 10

AC007615 AC007615 175691 bp DNA linear HUG-04 Jul. 2001  
 LOCUS Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT  
 DEFINITION  
 SOURCE AC007615  
 ACCESSION AC007615  
 VERSION AC007615.6 GI:14589428  
 KEYWORDS HUG: HUGS\_PHASE2; HUGS\_DRAFT; HUGS\_ACTIVEIN.  
 ORGANISM human.

REFERENCE 1 (bases 1 to 175691)  
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo:  
 1 (bases 1 to 175691)  
 TITLE DOE Joint Genome Institute.  
 JOURNAL Sequencing of Human Chromosome 16  
 AUTHORS 2 (bases 1 to 175691)  
 Bruce, D., Mandt, M., Poyet, N., Monk, C., Saunders, E., Robinson, D.,  
 Jones, M., Beckingham, J., Chastock, L., Thompson, S., Galloway, J.,  
 Bryant, J., Foster, J., McInock, J., Jamnare, J., White, S., Patino,  
 Campbell, C., Fawcett, J., Malhotra, M., Russel, M., Sutherland, R.,  
 McMurtry, K., Hancock, and Beaven, D.

TITLE Direct Submission  
 JOURNAL Submitted (20 MAY 1999) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M886, Los  
 Alamos, NM 87545, USA

COMMENT On Jul 3, 2001 this sequence version replaced at:14928451.  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.

Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.

NOTE: This is a 'working draft' sequence. It currently  
 consists of 1 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces  
 is believed to be correct as given, however the sizes  
 of the gaps between them are based on estimates that have  
 been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 175691: contig of 175691 bp in length.  
 location/Qualifiers  
 1. 175691

FEATURES  
 source  
 1. 175691  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-528K16"  
 BASE COUNT 42392 a 42125 c 45277 g 45846 t 1 others  
 ORIGIN

Query Match 9.0%; Score 187; DB 2; Length 175691;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-87;  
 Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1328 tactgtaaatcagatcattatcttttgaatcctgctgacacatattatattacac 1487  
 |||||||  
 Db 3166 TACTGTAATATATGCTAGATTTTGTGTTACCTGGCTGCACATATATATTAATATAC 31665



FEATURES  
Source

Location/Qualifiers  
1. 191496  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-501p17"

BASE COUNT 47031 a 47510 c 47005 g 45732 t 4218 others  
ORIGIN

Query Match: 9.0%; Score 187; DB 2; Length 191496;  
Best Local Similarity 100.0%; Pred. No. 47e-87;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1328 tacltgaataatcagctagatcttgcattgaatgctgacatatacctataaccca 1387  
|||||  
DB 26623 TACGTAAATCAATGCTAGATCTTTGCTTAACGCTGCACATATACCTTAATACCA 26682  
|||||  
DB 1388 ggaatcggagacaagaagcattctgcattgctctcctatgacatttacttaagctg 1447  
|||||  
DB 26683 GGATGGCGGACAAAGGCAATCTGCGATCTTCTCTCCATGAGCAATTTACTCAGCTG 26742  
|||||  
DB 1448 ccaagctgcttctacaccttcttcttcttcttcttcttcttcttcttcttcttctt 1507  
|||||  
DB 26743 CCAAGCTGCTCTTCTACACCTTTCTTTTACACACCAAGCAGCTTTTCAACCTGAA 26802  
|||||  
DB 1508 agaagat 1514  
|||||  
DB 26803 AGAAGCT 26809

RESULT 12  
AC009093 208008 bp DNA linear HTG 04-SEP-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-426c22, WORKING DRAFT  
DEFINITION  
SEQUENCE 1 ordered pieces.  
AC009093  
AC009093.8 GI:15426058  
HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens  
Human.  
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 208008)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 208008)  
DOE Joint Genome Institute.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 208008)  
DOE Joint Genome Institute.  
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
on Sep. 4, 2001 this sequence version replaced gi:13786700.

COMMENT  
Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
Genbank flat file format but are available as part  
of this entry's ASN.1 file.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
Genbank flat file format but are available as part  
of this entry's ASN.1 file.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
Genbank flat file format but are available as part  
of this entry's ASN.1 file.

NOTE: This is a "working draft" sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the bases  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 208008; contig of 208008 bp in length.  
Location/Qualifiers  
1. 208008  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-426c22"

BASE COUNT 49489 a 50452 c 53814 g 54252 t 1 others  
ORIGIN

Query Match: 9.0%; Score 187; DB 2; Length 208008;  
Best Local Similarity 100.0%; Pred. No. 4.7e-87;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1328 tacltgaataatcagctagatcttgcattgaatgctgacatatacctataaccca 1387  
|||||  
DB 28428 TACGTAAATCAATGCTAGATCTTTGCTTAACGCTGCACATATACCTTAATACCA 28547  
|||||  
DB 1448 ccaagctgcttctacaccttcttcttcttcttcttcttcttcttcttcttcttctt 1507  
|||||  
DB 28618 CCAAGCTGCTCTTCTACACCTTTCTTTTACACACCAAGCAGCTTTTCAACCTGAA 28667  
|||||  
DB 1508 agaagat 1514  
|||||  
DB 28668 AGAAGCT 28674

RESULT 13  
AC092375 173166 bp DNA linear RPI 02-NOV-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-645c24, complete sequence.  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens  
Human.  
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 173166)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 173166)  
DOE Joint Genome Institute.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 173166)  
DOE Joint Genome Institute.  
Submitted (02-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
on Nov. 2, 2001 this sequence version replaced gi:15667998.

COMMENT  
Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.

Sequence quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.



```

misc_feature /note="assembly_name:Contig27"
7718..9983
/note="assembly_name:Contig28"
misc_feature /note="assembly_name:Contig28"
10084..113594
/note="assembly_name:Contig29"
misc_feature /note="assembly_name:Contig29"
13695..18473
/note="assembly_name:Contig30"
misc_feature /note="assembly_name:Contig30"
18574..24956
/note="assembly_name:Contig31"
misc_feature /note="assembly_name:Contig31"
25057..34405
/note="assembly_name:Contig32"
misc_feature /note="assembly_name:Contig32"
34506..43064
/note="assembly_name:Contig33"
misc_feature /note="assembly_name:Contig33"
43165..50773
/note="assembly_name:Contig34"
misc_feature /note="assembly_name:Contig34"
50874..60610
/note="assembly_name:Contig35"
misc_feature /note="assembly_name:Contig35"
60711..69701
/note="assembly_name:Contig36"
misc_feature /note="assembly_name:Contig36"
69802..80031
/note="assembly_name:Contig37"
misc_feature /note="assembly_name:Contig37"
80132..93512
/note="assembly_name:Contig38"
clone_end:SP6
vector_side:left"
misc_feature /note="assembly_name:Contig39"
93613..111039
/note="assembly_name:Contig40"
misc_feature /note="assembly_name:Contig40"
11140..127223
/note="assembly_name:Contig41"
misc_feature /note="assembly_name:Contig41"
127324..149676
/note="assembly_name:Contig42"
misc_feature /note="assembly_name:Contig42"
149777..179150

```

BASE COUNT 51425 a 39220 c 37611 g 49085 t 1809 others

## Query Match

Best Local Similarity 8.5% Score 176: DB 2: Length 179150;  
Matches 176: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

QY 1339 catgcctagacatcttgaactgagcagacatatacttaataaacaagatctcga 1398
|||||
DB 72665 CATGCTAGATCTTTTCTTAACTGGCTGCACATACCTTAATVACCAHSAITTCGGAA 72606
|||||
QY 1339 caagagcatctgagatggtctctccatggagcatttaccacacctggcaagctat 1458
|||||
DB 72605 CAAGGAGATTTCTGGAGATTTGCTCTCCATGAGACCATTTTACTCAAGCTGCAAGCTTGT 72546
|||||
QY 1459 tctacacatttgaactgagcagacatcttgaagcggagaaacctgaataaagt 1514
|||||
DB 72545 TCTACACCTTGTGTTTACGACCAAGCAGCTTTTGGAGGAAAGCTCAAGACAGCT 72490
|||||

```

## RESULT 15

HFAP001549 202004 bp DNA linear PRI 28-JUL-1998  
LOCUS Human chromosome 16 HAC clone C19B7SK-A-270G1, complete sequence.  
DEFINITION AF001549  
ACCESSION AF001549.1 GI:3355302  
VERSION HTG.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 202004)  
AUTHORS Adams,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (29-APR-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
REFERENCE 2 (bases 1 to 202004)  
AUTHORS Adams,M.D., Lottus,B.J., Zhou,L., La Bombard,M., Kim,U.J. and  
Ventner,J.C.

TITLE Direct Submission  
JOURNAL Submitted (22-NOV-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
REFERENCE 3 (bases 1 to 202004)  
AUTHORS Adams,M.D. and Lottus,B.J.  
TITLE Direct Submission  
JOURNAL Submitted (28-JUL-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA

## COMMENT

On Jul 30, 1998 this sequence version replaced at:240964.  
HAC clone C19B7SK 270G1 is located in chromosome 16. Genes were  
identified by a combination of five methods: XGRAIL (available by  
anonymous ftp from arthur.epm.ornl.gov), Genzinder (available by  
anonymous ftp from coloww.washington.edu), GENSCAN (available  
using the e-mail server at genscanqumc.stanford.edu), Snatches  
of the EST database at 1108 (<http://www.fair.org/db/hw/hw.html>)  
and searches against a peptide database. Repeats were identified  
using RepeatMasker (Smith, A. and Green, P. unpublished,  
<http://ftp.jenome.washington.edu/rm/RepeatMasker.html>)  
Location:Quailifiers

## FEATURES

## Source

```

1..202004
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p12"
/contig="A-270G1"
join(46775..46941,47279..47384,48522..48578,54429..54518,
54625..53754,54203..54262,55202..55449)
/contig="A-270G1.2"
46775..55449
/contig="A-270G1.2"
join(47289..47384,48522..48578,54429..54518,54625..54702)
/contig="A-270G1.2"
/codon_start=1
/product="unknown gene product"
/protein_id="AAC2782.1"
/db_xref="GI:3355304"
/translation="MRALNDENFSPRKIVRGIVTPEVLKTKKGEINPEIKNG
LSDPKIKDQILLMLERSVMTLIDPFQILSLIKRIPMIINRQIVTVERIATLON
LVSAA"
54435..54535
/contig="A-270G1.2"
/db_xref="dbSTS:627479"
join(63313..63405,65204..65334,66207..66391,68184..68292,
68925..69073)
/contig="A-270G1.1"
63313..66073
/contig="A-270G1.1"
join(63403..63405,65204..65334,66207..66391,68184..68292,
68925..68958)
/contig="A-270G1.1"
/codon_start=1
/product="unknown gene product"
/protein_id="AAC2782.1"
/db_xref="GI:3355304"
/translation="MGFAKATLHLEKRLDIPSPNPAITLQAAATLISLTAAKATLSL
ITKPELIDLVNWLITLITLNSQISLKAFQVLAHGFYSNCAVYIVTVPKRLTSS
NLRKGLQVPSQINPERIVMSQINPLIKLIPSVNFALITNRRKICVQWM"
14745..174876
/db_xref="dbSTS:619499"
14872..185077
/db_xref="dbSTS:604856"

```

BASE COUNT 56977 a 45624 c 44447 g 54956 t

## ORIGIN

Query Match 8.5% Score 176: DB 2: Length 203604;  
Best Local Similarity 100.0% Pred No. 30-81;  
Matches 176: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

QY 1339 catgcctagacatcttgaactgagcagacatatacttaataaacaagatctcga 1398
|||||
DB 66318 CATGCTAGATCTTTTCTTAACTGGCTGCACATACCTTAATVACCAHSAITTCGGAA 66277
|||||

```







```

900  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 107
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
213  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 262
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
107  TactTactTactTactTactTactTactTactTactTactTactTactTact 123
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
263  AGAAATATATAAAATATATATATATATATATATATATATATATATATAT 412
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
124  THVVALVGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 140
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
413  AVGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDT 462
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
140  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 157
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
463  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 412
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
157  TactTactTactTactTactTactTactTactTactTactTactTactTact 173
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
414  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 461
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
174  ASPTGLASPTGLASPTGLASPTGLASPTGLASPTGLASPTGLASPTGLASPT 190
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
462  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 511
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
190  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 207
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
512  GTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 561
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
207  TactTactTactTactTactTactTactTactTactTactTactTactTact 223
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
562  AGAAATATATAAAATATATATATATATATATATATATATATATATATAT 611
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
224  CTSYTVVALHSASPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLY 240
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
612  TactTactTactTactTactTactTactTactTactTactTactTactTact 661
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
240  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 257
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
662  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 710
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
257  SVALSVALSVALSVALSVALSVALSVALSVALSVALSVALSVALSVALSVAL 273
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
711  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 760
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
274  CTSYTVVALHSASPTGLASPTGLASPTGLASPTGLASPTGLASPTGLASPT 290
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
761  TactTactTactTactTactTactTactTactTactTactTactTactTact 807
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
808  AGAAATATATAAAATATATATATATATATATATATATATATATATATAT 856
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
807  TactTactTactTactTactTactTactTactTactTactTactTactTact 823
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
857  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 906
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
824  TCTMTCTASASVALGYSYTVVALASPTGLASPTGLASPTGLASPTGLASPT 340
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
907  TactTactTactTactTactTactTactTactTactTactTactTactTact 953
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
940  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 956
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
954  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 997
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
956  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 973
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
998  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1046
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
973  TactTactTactTactTactTactTactTactTactTactTactTactTact 989
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1047  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1090
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

498  TactTactTactTactTactTactTactTactTactTactTactTactTact 406
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1091  GTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1146
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
406  TactTactTactTactTactTactTactTactTactTactTactTactTact 422
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1147  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1180
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
423  TCTMTCTASASVALGYSYTVVALASPTGLASPTGLASPTGLASPTGLASPT 439
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1181  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1229
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
439  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 456
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1240  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1270
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
456  TactTactTactTactTactTactTactTactTactTactTactTactTact 472
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1271  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1317
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
472  SVALSVALSVALSVALSVALSVALSVALSVALSVALSVALSVALSVALSVAL 489
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1318  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1365
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
489  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 505
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1366  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1405
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
505  TactTactTactTactTactTactTactTactTactTactTactTactTact 522
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1406  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1445
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
522  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 539
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1447  GTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1492
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
539  TactTactTactTactTactTactTactTactTactTactTactTactTact 555
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1493  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1539
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
555  GDTPLDAGVQVTLDPDAVTSQSSQSLQMLQTLVTLQVTHLYSASPPLDQ 572
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1545  TactTactTactTactTactTactTactTactTactTactTactTactTact 1589
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
572  TCTMTCTASASVALGYSYTVVALASPTGLASPTGLASPTGLASPTGLASPT 588
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1589  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1639
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
588  TactTactTactTactTactTactTactTactTactTactTactTactTact 605
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1640  CTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1688
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
605  TCTMTCTASASVALGYSYTVVALASPTGLASPTGLASPTGLASPTGLASPT 622
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1688  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1735
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
622  TactTactTactTactTactTactTactTactTactTactTactTactTact 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1736  TAAATATATAAAATATATATATATATATATATATATATATATATATAT 1784
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
638  GYSYTVVALASPTGLASPTGLASPTGLASPTGLASPTGLASPTGLASPT 643
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1785  GTRGQDTPVGGDAATCTGGTTCCTCTGCTATGCAATTCGAGAGAGATTCG 1799
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

seq\_name: rib\_pst2:86502871

seq\_date: not set

length: 86502871

DEFINITION: 6025504 top1 NIB\_moc6 61 Homo sapiens cDNA clone IMAGE:4559071 5'

FEATURES: mRNA sequence.

ACCESSION: 86502871

VERSION: 86502871.1

KEYWORDS: EST

size: 1446488

EST: 27 MAR 2001

SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 788)  
AUTHORS NIH-MGC <http://mgl.nhl.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [carphs@mail.nih.gov](mailto:carphs@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA library preparation: CLONTECH Laboratories, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
plate: LHCMI451 row: 1 column: 08  
High quality sequence stop: 699.  
FEATURES  
location/Qualifiers  
1..788  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4658071"  
/clone\_lib="NIH-MGC-61"  
/tissue\_type="embryonal carcinoma"  
/lab\_host="DM10B (T1 phage-resistant)"  
/note="Organ: testis; Vector: pMR-118 (clontech); Site-1: Still (ggcgcctgcgcgc); Site-2: Still (ggcgcctgcgcgc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCCTACGCGCCGCGCCGCGCATGCTG(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."  
BASE COUNT 204 a 178 c 176 g 228 t 2 others  
ORIGIN  
alignment\_scores:  
Quality: 1207.50 Length: 265  
Ratio: 4.792 Gaps: 5  
Percent similarity: 95.094 Percent Identity: 92.453  
alignment\_block:  
US-09-932-678-2 x HG502871 ...  
Align seq 1/1 to: BG502871 from: 1 to: 788  
302 ArgLeuAspGlnMetValHisProValAlaGluArgLeuAspGlnLeuMet 418  
|||||  
2 CGGCTCGACGACGATGTCATCTCTACCGACGCGCCGCGCATCTCAT 51  
318 TserLeuValIleuSerTyrMetLeuAspValCysTyrValAspGlyLeu 335  
|||||  
52 GCTTTGTTGTTTGTCTACATGAGGATGCTCTCTATGATGATGTAAGG 101  
335 a1AspAsnGlyLysGlnLysAspLeuTyrArgAspLeuIleuAsnLeu 351  
|||||  
102 TTGATTAAGGCAACAAAGATCTATATTCGACATGATATTAACATCTTT 151  
352 AspLysIleuLeuLeuProThrHisAlaSerCysHisValGlnPhePhe 368  
|||||  
151 GACAACATCTCTGTTCCACCCACCTCTCTGACATGACATCTTTTCAT 201  
368 PheTyrLeuGlySerPheLysLeuGlyPheAlaGlnAlaPheLeuGln 385  
|||||  
202 GTTATACCTCTGTAGTTTCAATTTGGATTCGACAGGCAATTTTTCGAC 251

385 TserLysLysIleuLeuAspProSerAspProValIleuLeuGln 401  
|||||  
252 ATCTCTGCAAAAATATTCAGAGATCCAGATATCTGCTCATPAAAGATG 401  
402 AlaAlaGlyAsnTyrIleGlySerPheLeuAlaAlaValAspPheLeu 418  
|||||  
302 GCTCTGCGAATATATATGAACCTTTTGGACAGCATTAATTATTC 351  
418 GlnLeuIleThrValLysSerCysLysAspLeuLeuValAsnLeuLeu 445  
|||||  
352 TCTATTACTGTAATATCATGCTAGATCTTTGCTTACTAGCTGACACA 401  
445 LeuTyrLeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAla 451  
|||||  
402 TATACCTTAATACCGAGATTCGGACAAAGCATTTTCAGATCTCT 451  
452 LeuHisGlyProPheTyrSerAlaCysGlnAlaValPheTyrIlePhe 468  
|||||  
452 CCGATGGACCATTTTACTTCAGCTCCACGCTGCTTCTACACCTTGT 501  
468 PheAlaHisLysGlnLeuLeuSerGlyValAspLeuLysGlyLeuGln 485  
|||||  
502 TTTTACGCAAAACACAGCTTTTACGCGAAGCTGAAGAAAGTTNCGAG 551  
485 TyrLeuGlnSerLysAspProValAlaValMetSerGlnLeuAspPro 501  
|||||  
552 ATCTTAAAGCTGAATTTGACGGGATGATATAGTCACTTAATCTC 601  
602 LeuLysLysLysLeuProSerValValAspPheAlaAlaIleTyrAs 518  
|||||  
602 GCAAGATTTCTGCGCAGACGCTTAATCTTTTTCGCAATCAACAA 650  
518 LysTyrGlnLeuValPheCysTyrThrIleGlnGluTyrAsnAsnAla 544  
|||||  
651 TTAAATGACAGCTCTTTTGTATACATATATGAGAGAGAAATAG 700  
544 GlnMetLeuProValIleArgSerThrAlaGlyLysAspSerVal 549  
|||||  
700 GCAAGATGTTGTCATCATTTGCACTACCTTGAGAGCAACATCAGC 750  
549 TcLcLcLcGcStLcAsnProLeuAspThrPhePhePro 561  
|||||  
751 GCAATCTGGCTAAACCCGCTGGAACGCTTTTCC 787  
seq\_name: gb\_est2:BF203684  
seq\_dont\_mention\_block:  
LOCUS BF203684 996 bp mRNA linear EST 06-NOV-2000  
DEFINITION 60186631f1 NIH-MGC\_17 Homo sapiens cDNA clone IMAGE:4094398 5'.  
ACCESSION EF203684  
VERSION BF203684.1 GI:11097270  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 996)  
AUTHORS NIH-MGC <http://mgl.nhl.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [carphs@mail.nih.gov](mailto:carphs@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA library preparation: Ling Hong/Robin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
plate: LHC9667 row: n column: 07  
High quality sequence stop: 684.  
FEATURES  
location/Qualifiers  
1..996  
SOURCE



```

1 MetAlaIaIaProIeuLeuHisThrArgLeuProGluYAspAlaIaIaIaSe 17
|||||
23 ATGGGCGGACCGCTGCTTACACAGCGCTTCCGGGAGATGCGCGCGCTTC 72
|||||
17 tSerSerAlaValIysIysIeuGlyAlaSerArgThrGlyIleSerAsn 34
|||||
73 GTCCTCTGCACTTAAAGAAAGCTGGAGGCTGGAGACTGGCATTTTCAATA 122
|||||
34 eIArgAlaIeuGluAsnAspPheAsnSerProProArgIysThrVal 50
|||||
123 TCGTGCAATTAGAAATGACCTTTTTCATTTCTCCCAAAAAAATTCGTT 172
|||||
51 ArgPheGlyGlyThrValThrGluValIeuLeuIysThrIysIysGly 67
|||||
173 CGATTGGGAACTGTACAGAAAGCTCTCTAAGTACAAAAAGGCTGA 222
|||||
67 uThrAsnAspPheGlyIeuLeuGlyAsnGlnIeuLeuAspProAspIleL 84
|||||
223 AACAAATGACTTGGAGTTTGAAAGAACCGACTGTAGATCAGACATAA 272
|||||
84 yAspAspGlnIleIleAsnTrpIeuLeuGlnPheArgSerSerIleMet 100
|||||
273 AGCATGACACACTCACTACACTGGTGTCTAAGATTCGCTTCTATCAG 322
|||||
101 tYrLeuThrIysAspPheGlyIeuLeuIleSerIleIleLeuArgLeuP 117
|||||
423 TACTTACAAAAGACTTGGACCACTTATCAGATATATATTAAGATTGCC 372
|||||
117 cTrpIeuAsnArgSerGlnThrValValGluGluIyTrpAlaIaPheLeu 134
|||||
373 TTGGTTAAATAGAACTCAACACTAGTGCAGAAAGTATTGGCTTTGTTG 422
|||||
134 IysAsnIeuValSerAlaGlnThrValPheLeuArgProGlyLeuSerMet 150
|||||
423 GTAACTTGTATCAGCAGACAGCTGTTTCTCAGACCGCTCTCCACCATG 472
|||||
151 tLeuIleSerHisPheValProProArgValIleIleCysGluIysPhe 167
|||||
473 ATTGCTTCCCATTTTGCTCTCCGAGTATCAATTAAGAAAGGATATGT 522
|||||
167 IAspValSerAspSerAspAspGluAspAspAsnIeuProAlaAsnPheA 184
|||||
523 ACATGTTTCAGATCTCATGATGAACATCAATATCTCTGCAAAATTGG 571
|||||
184 sPThrCysHisArgAlaLeuGlnIleIleAlaArgIyTrValProSerThr 200
|||||
572 ACACATGTCACAGACCTTCCAAATAAATACCAAAATATGTACCATGACA 621
|||||
201 ProTrpPheIeuMetProIleIeuValGluIysPheProPheValArg 216
|||||
622 CGTGTGCTTTCATGACCAATATNGTGGAAAAATTTTCATTTGTTGG 670
|||||
217 tYsSerGluArgThrIeuGlyGlyThrValHisAsnIeuLeuArgIle 232
|||||
671 AATTCACAGACAGACACTGAAATGTACCTTCAATATACCTAACGAAT 718
|||||
seq_name: gb_est2:BC428305
seq_documentation_block:
LOCUS BC428305 1037 bp mRNA linear EST 14-MAR-2001
DEFINITION 60249898.HPI NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4612605 5',
mRNA sequence.
ACCESSION BC428305
VERSION BC428305.1 GI:13334811
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1037)
AUTHORS NIH-MGC http://mgt.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

```

```

JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: raab@r-mail.nih.gov
Tissue procurement: Clontech Laboratories, Inc.
CDNA library preparation: Clontech Laboratories, Inc.
CDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1361 row: e column: 22
High quality sequence stop: 637.
FEATURES
SOURCE
1..1037
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4612605"
/clone_lib="NIH MGC 75"
/lab_host="PH109 (T1 phage-resistant)"
/vector="Origam: Kidney: Vector: pDNR-LTR (Clontech). Site 1
and 3 adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGGCAATATGGC-3' and 4' adaptor sequence:
5'-ATTCTAGAGGCGCGAGCGCGGACAGT-d(10)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC library."
BASE COUNT 292 A 218 C 203 G 204 T
ORIGIN
aligner: t_block:
Quality: 1100.50 Length: 400
Ratio: 4.031 Gaps: 14
Percent similarity: 91.000 Percent identity: 84.333
aligner: t_block:
US-09-32-678-2 x BC428305
Align seq 1/1 to: BC428305 from: 1 to: 1037
1 MetAlaIaIaProIeuLeuHisThrArgLeuProGluYAspAlaIaIaIaSe 17
|||||
24 ATGGGCGGACCGCTGCTTACACAGCGCTTCCGGGAGATGCGCGCGCTTC 71
|||||
17 tSerSerAlaValIysIysIeuGlyAlaSerArgThrGlyIleSerAsn 34
|||||
72 GTCCTCTGCACTTAAAGAAAGCTGGAGGCTGGAGACTGGCATTTTCAATA 121
|||||
34 eIArgAlaIeuGluAsnAspPheAsnSerProProArgIysThrVal 50
|||||
172 TCGTGCAATTAGAAATGACCTTTTTCATTTCTCCCAAAAAAATTCGTT 171
|||||
51 ArgPheGlyGlyThrValThrGluValIeuLeuIysThrIysIysGly 67
|||||
173 CGATTGGGAACTGTACAGAAAGCTCTCTAAGTACAAAAAGGCTGA 221
|||||
67 uThrAsnAspPheGlyIeuLeuGlyAsnGlnIeuLeuAspProAspIleL 84
|||||
223 AACAAATGACTTGGAGTTTGAAAGAACCGACTGTAGATCAGACATAA 271
|||||
84 yAspAspGlnIleIleAsnTrpIeuLeuGlnPheArgSerSerIleMet 100
|||||
273 AGCATGACACACTCACTACACTGGTGTCTAAGATTCGCTTCTATCAG 321
|||||
101 tYrLeuThrIysAspPheGlyIeuLeuIleSerIleIleLeuArgLeuP 117
|||||
423 TACTTACAAAAGACTTGGACCACTTATCAGATATATATTAAGATTGCC 371
|||||
117 cTrpIeuAsnArgSerGlnThrValValGluGluIyTrpAlaIaPheLeu 134
|||||

```





474 . 1

$\text{N}(\text{C}_6\text{H}_5)_3$	207	215	217
$\text{N}(\text{CH}_3)_3$	207	215	217

[illegible]

Albuquerque, NM 87101 to: 828

[illegible]

204 GCGAGGATGCTAAATTAATCTGGCTATCAATCAAGGCTAAAGGCTGGCTGCTGCAATTTATA 252

424 SGTCTYSLEDAIPLEQLEQVLAASUTPLLEHLISLETYLAQASNAAGCI 440

440 **inspectorily**ThruysAtoHwysAsPValAlal,eunHSLyptroPher 457  
|||||

404 A C T A C A C C T C C A A A C C C T C T C T C T T T A C A C C T T T C T T T A G A C A C C A A G C C A G 452

4153 GGTCTTGGATCTTCAAAACCTCAAAAGAAAGCTTTTCTCACTATCTTCAGAGCTCTCGAA 502

5.0.7  $\text{[O}^{\text{S}}\text{O}^{\text{V}}\text{Al}^{\text{I}}\text{V}^{\text{I}}\text{As}^{\text{I}}\text{m}^{\text{I}}\text{b}^{\text{I}}\text{h}^{\text{I}}\text{O}^{\text{I}}\text{Al}^{\text{I}}\text{Al}^{\text{I}}\text{Al}^{\text{I}}\text{e}^{\text{I}}\text{b}^{\text{I}}\text{rAs}^{\text{I}}\text{O}^{\text{I}}\text{I}^{\text{I}}\text{S}^{\text{I}}\text{V}^{\text{I}}\text{V}^{\text{I}}\text{I}^{\text{I}}\text{e}^{\text{I}}\text{V}^{\text{I}}\text{Al}^{\text{I}}\text{5.24}$

[illegible]

604 1011361ACAAATATCATTCGAGAGAGAAATATCGCGAGATGCTGCTACCTGAT 612

652 1AC2AC1A07AC01C6ACBACAC10AC16AC0A07C160AC0A0A0C00AC000702

Asp390104Y1 . . . . . 673

547 SPOT: SOLAR ACTIVITY INDEX IN 590  
11 11111  
11111

documentation\_block:

SS1 N	RI464721	01:1525364
SS1 N	RI464721.1	

Eukaryotic: Multicellular: Choanoflag: Ciliates: Vertebrates: Mammalia: Eutheria: Primates: Carnivora: Hominoidea

Unpublished (1959)  
Contact: Robert S. Lausberg, 4410

Yoshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.C. Consortium

```

rowid image_id row
plate_1A11676 row_b column_12
http://image.llnl.gov

```

for *organism* "Homosapiens"  
with word "exod-194444"

```

/ addcost = 0.001;
/mult = "Jordan"; test is: Vector = pluesesr1ptk
pluesesr1k; test is: Vector = pluesesr1k; test is:
pluesesr1k; test is: Vector = pluesesr1k; test is:

```

rectified and collected in a primary in  
for fuel - engines and engines used

2009	178	144	2009
178	144	2009	178
144	2009	178	144
2009	178	144	2009

Quality:	989.50	Length:	234
Kat 10:	4.477	Caps:	4







Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
<http://image.lnl.gov>

Plate: LHAM0247 row: 9 column: 10  
 High quality sequence stop: 665.

## FEATURES

Location/Qualifiers

1..990

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:445009"

/clone\_lib="NIH\_MGC\_89"

/issue\_type="hypermethylation, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: Kidney; Vector: PCMV-SPORE; Site: 1; Not1:

Site\_2: SalI; Cloned unidirectionally: oligo-dT primed.

Average insert size 1.3 kb. Library enriched for

full-length clones and constructed by Life Technologies

Note: this is a NIH\_MGC library."

BASE COUNT 304 a 275 c 219 g 197 t

ORIGIN

alignment\_scores:  
 Quality: 945.00 Length: 187  
 Ratio: 5.109 Gaps: 1  
 Percent Similarity: 97.861 Percent Identity: 97.326

alignment\_block:

US-09-932-678-2 x BG167061 ..

Align seq 1/1 to: BG167061 from: 1 to: 990

```

466 ThrPheValPheArgHisLysClnLeuLeuSerGlyAsnLeuysGlnG 482
|||||
2 ACCTTGTTGTTTACGACACAGAGCTTTGACCGAAGCTGAAAGAGC 51
|||||
482 yLeuGlnTyrLeuGlnSerLeuAsnPheGluAqLLeValMetSerGln 499
|||||
52 TTTCAGATCTTCAGACGTGAAATTTGACCGGATAGTATACGCCACG 101
|||||
499 eu.AsnProLeuLysLecysLeuProSerValValAsnPhePheLaa 515
|||||
102 TAAAGCCCTGACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 151
|||||
515 alLeThrAsnLysTyrGlnLeuValPheCysTyrThrLleLeGluArg 542
|||||
152 ATTCACAAATAGTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 201
|||||
532 snAsnArgLmetLeuProValLleArgSerThrLacLysLysPser 548
|||||
202 ACAATGCCAGATGCTGCCATCATTAAGACATACCCCTGAGAGAGCTCA 251
|||||
549 ValGlnLleCysThrAsnProLeuAspThrPhePheProPheAspPro 565
|||||
252 GTCCACATCTGTCACAAACCCCTGACACCTTCTCCCTTTGATCTG 301
|||||
565 sValLeuLysArgSerLysLysPheLleAspProLleTyrGlnValArg 582
|||||
302 TGTGCTGAGAGAGCTCAAGAAATTCATTGATTCATTACAGTGTGG 351
|||||
582 LuAspMetSerAlaGlnLleLeuGlnGluLlePheLysLysProMetLys 598
|||||
352 AAACATGATGCTGCTGAGAGAGCTTCAAGAAACCCATGAGAAAG 401
|||||
599 AspLleValGluAspGluAspAspAspPheLeuLysGlyGluValProG 615
|||||
402 GATCATGTGAGAGATGAAGATGATGATCTTGTAAAGAGGAGATGCTG 451
|||||
615 uAsnAspThrValLleGlyLleThrProSerSerPheAspThrHisPhe 632

```

|||||

472 GAAATATACCTTATGATGATGATGATGATGATGATGATGATGATG 501

634: 138PheProSerSerValLysSerProProValLeuLeuMetSerPro 648

|||||

502 GAAGCTCTCAAGATGATGATGATGATGATGATGATGATGATGATG 541

649 SerProLeu 651

|||||

552 ATTCCTCTC 560

seq\_name: qb\_est2:BG297178

seq\_documentation\_block:

LOCUS BG297178 1045 bp mRNA linear EST 21 FEB 2001

DEFINITION 602395174F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:450694 5'

VERSION BG297178

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1045)

AUTHORS NIH-MGC <http://imgc.net.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cygares@mail.nih.gov](mailto:cygares@mail.nih.gov)

Tissue Procurement: The Coriell Laboratory

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:

<http://image.lnl.gov>

Plate: LHAM0342 row: d column: 19

High quality sequence stop: 667.

Location/Qualifiers

1..1045

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:450694"

/clone\_lib="NIH\_MGC\_94"

/issue\_type="Telom"

/lab\_host="DH10B (phage resistant)"

/note="Organ: eye; Vector: PCMV-SPORE; Site: 1; Not1:

Site\_2: SalI; Cloned unidirectionally: oligo-dT primed.

Average insert size 1.3 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC library."

BASE COUNT 303 a 225 c 273 g 243 t 1 others

ORIGIN

Align seq 1/1 to: BG297178 from: 1 to: 1045

```

alignment_scores:  

  Quality: 928.50 Length: 346  

  Ratio: 4.613 Gaps: 10  

  Percent Similarity: 76.488 Percent Identity: 64.881

```

alignment\_block:

US-09-932-678-2 x BG297178 ..

Align seq 1/1 to: BG297178 from: 1 to: 1045

```

1 MetAlaAlaProLeuLeuHisThrArgLeuProGlyAspAlaAlaAsp 17
|||||
2c ATGGGAGGCGGCGGTGCTTCAACGCGCTTGTGGGCGATGTGAGGAG 76
|||||
17 pHisSerAlaValLysLysLeuGlnValAspSerThrGlyLysSerAs 44
|||||
76 GCGGCTCTGCGACGCTCTGAGTGGCGAGCGTGGGCTTTTGTGA 122

```



```

303 TTGAGCACTTATCACTATATATATAGATTCCCTGCTGATAGAG 352
122 GGTGTTTValValGluGluTyrLeuAlaPheGluGlyAspLeuValSerA 139
353 TCAACAGAGTATGAAAGAGATATGGCTTTCTGATATCTGTATCATGAG 402
139 GAGTATTTValPheLeuGATPrGlyLeuSerMetLeuAlaSerHisPhe 155
403 CACAGACTGTTTCTGAGACCGCTGCTACGATCATTCCTTCATTTT 452
156 ValProProAlaValIleIleLysGluGlyAspValAspValSerAsp 172
454 GTCCTCCCGGAGTATGATTAAAGCAAGCGATGATGATTTGATGAT 502
172 GATAPAPGluAspAspAspLeuProAlaAspPheAspPheGlyHisArg 188
503 CTCATGATGACAGATGATATCTTCTGCANATTTCACACATGTCACAGA 551
189 AlaLeuGlnIle 192
552 GCCCGCAATA 563
seq_name: qb_esc2:HG295610
seq_documentation_block:
LOCUS HG295610 956 bp mRNA linear EST 21-FEB-2001
DEFINITION 60239292F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4504751 5',
mRNA sequence.
ACCESSION HG295610
VERSION HG295610.1 GI:13057417
KEYWORDS EST.
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 956)
NIH-MGC http://mgi.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1995)
COMMENT Contract: Robert Strausberg, Ph.D.
Email: cga@bimail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Invitrogen, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10377 row: q column: 24
High quality sequence stop: 660.
location/qualifiers
1..956
/organism="Mus musculus"
/adb_xref="taxon:10090"
/clone="IMAGE:4504751"
/clone_1lb="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."
BASE COUNT 266 a 204 c 239 g 247 t
ORIGIN
alignment_scores:
Quality: 898.50 Length: 277
Ratio: 3.682 Gaps: 8
Percent Similarity: 88.087 Percent Identity: 71.841
alignment_block:

```

```

085-09-932-678-2 x HG295610
Align seq 1/1 to: HG295610 from: 1 to: 956
147 CysLeuSerMetLeuAlaSerHisPheValProProAlaValIleLeu 164
4 TGTCTGACATGATTCGGCTCATTTCTACCTGCTCCAGATATGTCAA 51
163 GCTGCTGAspValAspValSerAspSerAspPheGluAspAspLeu 179
52 GCAAGTGGCATGATGATGATTCAGATTCGATGACGAGCAATGACATCA 101
180 ProAlaSerPheAspPheCysHisArgAlaLeuGlnIleLeuAlaArg 196
102 TCTGCGCATTTTGCACATGTCACACAGCCCTTGCAAAATATACAGAT 151
156 ValProProThrProThrPheLeuMetProIleLeuValGluLysPhe 212
152 ATCTCCATGACACGATGATTTTAAATGCCAATCTGCTGGAGAAATTC 201
213 ProPheValArgLysSerGluArgThrLeuGluGlyValIleHisAsn 229
202 CGGTTGTAGAGATGCCAGAGAAATTCGATGCTATGCTTCATACAT 250
229 GLeuArgIleSerValIlePheProThrLeuArgHisGlnIleLeuGlu 246
251 ATTAGAGATAGTTTAAATTCGCAACTTCAGCCGTGAAATTCGGAGC 300
246 GLeuLeuLeuLysLeuLeuLysLeuAspValAspAlaSerArgHisG 252
301 TGTATTGAAAGCTACTCACTTACATGATGATGATGATGATGATGAT 350
263 TLeuLysPheLeuGlnIleAlaThrIleThrGlyGlyIleThrAsp 279
451 ATTAGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400
279 TThrGlyGlyLeuPheAsnMetAspLeuAspLeuThrGlnHisArg 296
401 CACAGAGAGAGCTGTTTATATGATGATGATGATGATGATGATGATG 450
256 TrpLysAlaGlyProGluArgLeuAspGlnMetValHisProValAla 312
451 AAGAGGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 500
313 ArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysAspVal 329
501 CAGCTGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
325 TrpValValAspGlyLysValAspAsnGlyLysThrLysAspLeuTyr 345
571 CGGTGTCACGCTTAAATTTAAATTAACATTAACAAAGAGATTAAAC 600
345 SerLeuLeuAspIlePheAspLysLeuLeuLeuProIleHisAlaSer 361
601 ATCTGATATGATATTTGATGATTAAGCTCTCTTTTGCCACACATGCT 650
361 TCTGTTTValGlnPhePheMetPheTyrLeuGlySerPheLysLeuG 377
651 CTGCGATGATGAGCTGCTTTCATGCTTTCTTCCGCGAGCGCTTCAAG 700
377 TProAlaGlnAlaPheLeuGlnHisLeuThrLysLysLeuAlaPro 394
701 GATGCGCAGAGAGCTATTTCGAAACATCTCTGACCAAGAGTCGACAT 750
394 SerAsnProAlaIleLeuArgGlnAlaAlaGlyAsnTyrIleGlySer 410
751 ATAAAG...CGATTAATCAGAGAG...CGATGAAATATTTCGTAAGCT 794
411 GLeuAlaArgAlaLysPheIlePro 418
795 T...GCCAGACATTAATATATTC 816

```



[illegible]

7. SASSNILL





```

305 GluMetVal.....HisProValAlaGluArgLeuAsp.....11 316
|||||.....|
6324 CAAAGTTTGGATGACACACATTTGGACACGACGATACACACATGGAAAG 6373
316 GluMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspG 333
|||||.....|
6374 GTTATCAACATATATGTTAACCCACGACAGTTG.....C 6411
333 LysValAspAsnGlySerThrLysAspLeuTyrArgAspLeuLeuAsn 449
|||||.....|
6412 CGCATATGATGCTGGTTCAGCTCTTATCTCTTAACCAATATCAGTA 6461
350 LLeuAspLysLeuLeu.....LeuProTh 458
|||||.....|
6462 GTTAAGCAACATATTCACATTCGCTCTTACTGTATATATTATAC 6511
358 ThrAlaSerCysHisValGln..... 465
|||||.....|
6512 GTACCGAATATGATATAGAACTACAGATCTGCTGAAGGTTCAG 6561
365 ..... 365
6562 ATAAAGTACATTTGTTCTGTCGACAGGTATATTCCTGCTTG 6611
366 PhePheMetPheTyrLeuGlySerPheLys.....LeuGlyPheAlaG 480
|||||.....|
6612 TATATACACATATATCTCTGATCTCAACCAATTGCTTACGTTGTTCA 6661
380 ValPheLeuGlnHisLeuTyrLysLysLeuGlnAspProSerAspPro 397
|||||.....|
6662 GCGACAGCTAAGATTAATTCTGCGAAGATACGACATTTCTCAAGTATA 6711
397 LeuLeuLeuGlnAlaAlaGlyAspTyrTrpGlySerPheLeuAlaArg 413
|||||.....|
6712 GCTTCCTTACGACAAATGGATTTAGCATATCTCATTTGTTAGGCAAA 6761
414 AlaLysPheLeuProLeuLeuThrValLysSerCysLeuAspLeuVal 430
|||||.....|
6762 TTGACAGACCT.....TTACGTTACGCTGATTTGATTAAT 6799
430 LAsnTrpLeuHisLeuTyrLeuAsnAsnGlnAspSerGlyThrLysAlaP 447
|||||.....|
6800 CCACTTAAACAT..... 6812
447 HeCysAspValAlaLeuHisLysProPheTyrSerAlaCysAlaVal 463
6812 ..... 6812
464 PheTyrThrPheValPheArgHisLysGlnLeuLeuSerGlyAsnLeu 480
|||||.....|
6813 .....CAATTCACATCA.....CTCA 6829
480 SerLeuLysLeuGlnTyrLeuGlnSerLeu.....AsnPheGluArgLe 494
|||||.....|
6830 GGAGGCTTATTTGCTTAACATCATTCATTCATTTTTCAGAACCT 6879
495 .....ValMetSerGlnLeuAsnProLeu 502
6880 ATGATAGCATATGACCTTTGCTGCTTATAGACAGATTTCTGTAAG 6929
503 LysLysCysLeuProSerValValAsnPhePheAlaAlaLeuThrAsn 519
|||||.....|
6930 GCATACAGCTGACGATGATTCATTCATTCGCTGCTGCTGCTGCTGCT 6979
519 S.....TyrGlnLeuValPheCysTyrThrLeuLeuArgAsnAsn 534
|||||.....|
6980 ACTGTGATACAAAGTTCTTTGGATTTCTGAAGCTTTTGGAG...ATATTA 7026
534 TrpGlnMetLeuProValLeuArgSerThrAlaGlyLysSerValGln 550
|||||.....|
7027 AGTTCTTAATTAAGCTTCTGTCGACATCTGAAAGACAGACATTTGAA 7076
551 LLeuGlyThrAsnProLeuAspThrPhePheProPheAspProCysValLe 567

```

```

7077 GTTATGTCATGAGAA.....61 7094
|||||.....|
567 ULysArgSerLysLysPheLeuAspProTyrGlnValTrpGluAspM 984
|||||.....|
7094 TSCAAGACAGACACATATGATGACAGATTTTTCAGCTTATATCTAAA 7146
584 ULysArgLeuGlnLeuGlnPheLysLysTrpGlnLysLysAspL 600
|||||.....|
7144 GAGCAACGAGAGAAATGAGAGCTTTTTCATATCAATA..... 7181
604 ValGluAspLysAspAspPheLeuLysGlyGluValProGlnAsnAs 617
|||||.....|
7182 .....CATGAATTAAGCAATAAGTACTTGCAGATTCAGCTTGA 7225
617 P.....PheValLLeuGlyLeu..... 622
|||||.....|
7226 TGTCAATTCAGATGTCATGTCAGATTCGAGATTCGAGATTCAGATTC 7275
624 .....TheProSerPheAspThrHis 630
|||||.....|
7276 ACAAATTAACATCAATCAAAATGACCTGCTGCTGCTGCTGCTGCT 7322
seq_name: /seq2_6/prodata/1/lna/5R_CDMR.seq:US-08-980-427-1
seq_documentation_block:
? Sequence 1, Application US/08680427
? Patient No. 5854321
? GENERAL INFORMATION:
? APPLICANT: Staskiewicz, Brian S., Oldroyd, Giles Edward.
? ALIQUANT: Salmeron, John M., Romeros, Carlos
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
? NUMBER OF SEQUENCES: 15
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Klarcquist, Sparkman Campbell, Lejah S
? STREET: One World Trade Center
? STREET: 121 S.W. Salmon Street
? STREET: Suite 1600
? CITY: Portland
? STATE: Oregon
? COUNTRY: United States of America
? ZIP: 97204
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Disk, 3-1/2 inch
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: MS DOS
? SOFTWARE: Mofdirect 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/680,427
? FILING DATE: July 11, 1996
? CLASSIFICATION: 800
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/410,912
? FILING DATE: September 22, 1994
? CLASSIFICATION: 800
? APPLICATION NUMBER: 08/227,460
? FILING DATE: April 13, 1994
? CLASSIFICATION: 800
? ATTORNEY/AGENT INFORMATION:
? NAME: Dow, Alan, E.
? REGISTRATION NUMBER: 35,124
? REFERENCE/DOCKET NUMBER: 5151-45078
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (503) 226-7391
? TELEFAX: (503) 226-9446
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5475 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double stranded
? ORIENTATION: linear

```







```

Ratio: 0.387      gaps: 25
Percent Similarity: 45.455      Percent Identity: 19.755
alignment block:
US-09-9432-678-2 x US-09-080-897-3      ..
Align seq 1/1 to: US-09-080-897-3      from: 1 to: 4378

```

```

1230  |CAGTTTGGTGAAGTCTTTTCATCATATTATTAAACAGAGTCAAAAGATTTCA 1279
201  |.....PRTTTPheLeuMetProIleLeuValGluLeuSphProIle 214
1260  |AGCCAGACGCAACACCTCCCTGCT.....ATCTGGACATGCTCTGTG 1428
215  |ValAlaGlySerGluMetThrLeu.....Al 223
1424  |GTCCGAATATATATTAACAGACGCGCAACAGACATATAAATATATGAAA 1478
222  |cGstyValHisAspLeuLeuAlaIleSerValTyrProGlnIleGlu 240
1374  |AGTGTTTCTCAAAAGTCTTCACAAATAAGAACTGATCTGCACTTCA 1428
240  |GAGAGGAGTLeuLeuGlnIleIleGluIleLeuLeuLysLeuAspVal 256
1424  |ATGTGCACACGCTGAGATTCATATATGACATG.....GTGACATCA 1467
253  |AsnAlaSerArgGlnGlyIleIleGluAspAlaGluGluAlaAlaGlnThr 273
1466  |ATGATTGATTAACAAAGGTGGAAAAATTTGAGTCAAAATCAAACTCAAGT 1517
27  |TysTyrGlyThrAspSer.....ProGln 282
1516  |GCAAAAAAGCTGATATTCACAAATTAACAGCTGCGCAGATTAACAGAG 1567
242  |TyrLeuPheMetSerAspThrAsp.....GluGluThrGlnGlnSerGluThr 297
1576  |AAATGAAAAAAATGCAAAATGACATTCAGCAGCAAACTTACAGATCTTCA 1617
296  |AlaGlyTyrGlnMetIleLeuAspGlnMetValHisProValAlaGluAla 314
1614  |GAGAAAAAGATAGAGTATGATTCGTAAACAGACATCAATTCAGTCAAAA 1667
314  |GAspIleGluMetSerIleValLeuSerTyrMetLysAspValGstyVal 341
347  |U.....IleAsnIlePheAspIleLysLeuIleAspGlnGlnHisA 360
1766  |GTAGTTTGTCACTTCTCTTCTCACTGACGTGCTGCTGCTGCTGCTGCT 1817
360  |AspGstyHisValGlnPhe.....PheMetPheTyrLeuGstySerPheLys 375
1816  |TCTGTGCTGTGACCTGCTGACCTGCTATATGACGCTGCTGCTGCTGCTG 1867
376  |LeuGlyPheAlaGlnAlaPheLeuGlnHisLeuThrLysLysLeuHis 392
1816  |CTCTCTCTCTGAGAGTGGTGGCTGACAT..... 1867
392  |ProGstAsnProAlaIleLeuArgGlnAlaAlaGlyAsnTyrIleGln 409
1897  |..... 1897
409  |SerPheGluAlaArgAlaLysPheIleProGlnIleThrValLysSerGst 425
1895  |.....CCCTCTCTC.....GCT 1909
426  |LeuAspLeuLeuValAsnTyrPheGlnIleIleTyrLeuAsnAsnGlnAsp 442
1910  |CTCCAGTACTATCT.....CTCCAGTCTCTCT..... 1942
442  |TyrThrLysAlaPheGstAspValAlaLeuHisGlyProIleThrSerG 459
1943  |.....TACCTGAGAGTGGCTTGATAC.....CCCTGCTGCTGCTCA 1976
456  |ACAGGAlaAlaVal 463

```



```

257 ASDAlaSerArgGlnGlyIleGluAspAlaGluGluThrAlaThrGlnTh 273
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1468 ATGATTGATAAACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1517
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
273 cGcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 282
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1518 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1567
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
282 LyleuPheAsnMetAspGluAsp...GluGluThrGlnGlnGlnGlnGln 297
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1568 AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
298 ATAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 314
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1618 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1667
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
314 uAspIleLeuMetLeuLeuValLeuSerTyrMetLysAspValGlyTyr 331
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1668 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1717
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
331 aLAspGlyValValAspAsnGlyLysThrLysAspGlnGlnGlnGln 347
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1718 TGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1767
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
347 u.....LeasnIlePheAspLysLeuLeuLeuProThrHisA 360
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1768 GTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1817
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
360 LAspGlyHisValGlnPhe...PheMetPheTyrLeuLysSerPheLys 375
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1818 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1867
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
376 LeuGlyPheAlaGlnAlaPheLeuGlnHisLeuTyrLysLysLeuGln 492
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1868 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1897
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
392 ProSerAspProAlaIleLeuArgGlnAlaAlaGlyAsnTyrIleGly 409
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1897 ..... 1897
409 erPheLeuAlaArgAlaLysPheIleProLeuIleThrValLysSerG 425
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1898 .....CCCTCCCTCC.....TGC 1909
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
426 LeuAspLeuLeuValAsnTyrPheGlnHisTyrIleLeuAsnGlnAsp 442
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1910 CTCACAGTACTGTATGTC...CTCCACCTGCTGCTT..... 1942
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
442 rGlyThrLysAlaPheCysAspValAlaLeuHisGlyProPheTyrSer 459
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1944 .....TACCTGGAGAGTGTGTGTATAC...CCCTCCCTCCCTGCA 1976
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
459 LAcysGlnAlaVal 463
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1977 GTTCCCTGGCAGTGTG 1990
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
seq_name: /cqu2_6/prodata/1/lnu/6A_Comb.seq:us-08-899-595-2
seq_documentation_block:
: Sequence 2, Application US/0889595
: Patent No. 6111072
: GENERAL INFORMATION:
: APPLICANT: Natumaya, Shuh
: APPLICANT: Takahashi, No. 6111072nak1
: TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 4000 K Street, N.W., Suite 500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA

```

```

ZIP: 20007-5139
: Q. MEDIUM READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patcuth Release #1.0, Version #1.40
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/899,595
: FILING DATE: 24-JUL-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 8-242701
: FILING DATE: 26-AUG-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 9-90170
: FILING DATE: 25-MAR-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Stephen A. Bent
: REGISTRATION NUMBER: 29,768
: REFERENCE/WORK KEY NUMBER: 049441/0112
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5400
: TELEFAX: (202)672-5399
: INDEX: 904146
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4399 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: M. FEATURE TYPE: cDNA to mRNA
: ORIGINIAL SOURCE:
: ORGANISM: Mouse
: FLATOME:
: NAME/KEY: CDS
: LOCATION: 94..4858
: US-08-899-595-2

alignment_scores:
      quality: 100.50      length: 572
      ratio: 0.387      gaps: 26
      percent similarity: 45.455      percent identity: 19.755

alignment_block:
US-09-932-678-2 x US-08-899-595-2 ...
Align seq 1/1 to: US-08-899-595-2 from: 1 to: 4399

1 MetAlaAlaProLeuLeuHisThrArgLeuProGly.....AspAlaAl 15
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
438 AUGGTGTGCATATATCTGCACACTTCCAGAGTGCATATACATAGAGAA 485
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
15 GAAsrSerSerAlaValLysLysLeuGlyAlaSerArgThrGlyGln 42
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
486 GAGCTGTAGCTGTGCATATATATATATATATATATATATATATATAT 545
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2 erAsnMet.....ArgAlaLeuGlnAsnAspPhe 42
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
536 GAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
4 AsnSerProProArgLysThrValArgPheGly.....GlyTh 55
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
586 AATACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 635
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
55 ValThrGlnValLeu.....LeuLysTyrLysLysGlyGlnThr 68
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
636 CTAATAGCAGACATCTCAAGAGACATCATAGAGAGAGAGAGAGAGAG 685
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
69 .....AspAspPheGlnLeuLeuLysAsnGlnGlnGlnGln 79
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
696 GAACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```





Ratio: 0.381 Gaps: 21  
Percent Similarity: 46.691 Percent Identity: 19.499

alignment\_block:  
US-09-932-678-2 x US-08-973-462-2 ..

Align seq 1/1 to: US-08-973-462-2 from: 1 to: 5361

```

41 11eSerMetLeuAlaLeuGluAspPhePheAsnSerProIle 47
   :::::::::::::: ::::::::::::::::::::
3949 GATCAGATTAAAGATCTTCAGAGAAATATATTAAGAACTAAAGA 3998
47 9.....LysT 49
   :
3999 AATCAAAAGACTGAAAGTAAATTTCAGACATTAATAAGATTAAAAA 4048
49 11eValArgPheGlyCylThrValThrGluValLeuLysTyrLys 65
   :::::::::::::: ::::::::::::::::::::
4049 CATTTGAA.....ACAGATATTTACAGAGAAAAAAGAA 4083
66 GlyCylThrAsnAspPheGluLeuLeuLysAsnGluLeuAspProAs 82
   :::::::::::::: ::::::::::::::::::::
4084 ATAGAAAAAGATCATTTTGAAAAATTCAGAGAAAGAGCT.....GAGA 4127
82 pTLeuLysAspAspGluIleLeuAsnTyrLeuGluPheArgSer..... 97
   :::::::::::::: ::::::::::::::::::::
4128 AATAAAAAGATCTGAA...GCAGATATTTAAAGAGATATCTTCATTAC 4174
98 .....Ser11eMetTyrLeuThrLysAsp 105
   :
4175 AAGTGAAGAGAAAAAAATTTAGAGAAAGTACAGCAATTAAGAGAGAG 4224
106 pPheGluIleLeuSer11eLeuLeuArgLeuProTyrLeuAsnArgSe 122
   :::::::::::::: ::::::::::::::::::::
4225 GTAGAAACATATATTAAGTGGTGGATGCGCATATAAAGCTTTGAAAGAA 4274
122 pGluThrValValGluIleTyrLeuAlaPheLeuGlyAsnLeuValSer 139
   :::::::::::::: ::::::::::::::::::::
4275 TGAT.....TTAGAGAGAGTACATGATTTAAAGGAGATATA..... 4311
139 1aGluThrValPheLeuArgProCysLeuSerMet11eAlaSerHisPhe 155
   :::::::::::::: ::::::::::::::::::::
4312 .....TTAGACATGTTA..... 4323
156 ValProProArgVal11e1eCysGluGlyAspValAspValSerAspSe 172
   :::::::::::::: ::::::::::::::::::::
4324 .....AAGGAGATATGGAATTCAGCGGATAT 4349
172 AspAspGluAsp.....AspAsnLeuP 180
   :::::::::::::: ::::::::::::::::::::
4350 GGAATAGAGAAAGTTAGAGAGATGAACAAAGAACTGGAGAGAGAGCTG 4399
180 pGluAlaAsnPheAspThrCysHisArgAlaLeuGlu11e1eAlaArgTyr 196
   :::::::::::::: ::::::::::::::::::::
4400 AATCGTTAAAGATCTTTATCTAGTGCATTAGCGATGATGAGAGACAA 4449
197 ValProSerThrProTyrPheLeuMetPro11eLeuValGluIleLysPheP 213
   :::::::::::::: ::::::::::::::::::::
4450 ATGCAAAATCAAAAAAGCTCAAGACCTAAGCTGAGAGCAATATATATT 4499
213 pPheValArgLysSerGluAlaGluThrLeuGluCysTyrValHisAsnLeuL 230
   :::::::::::::: ::::::::::::::::::::
4500 AAAAGAGAGCGTTAAAGAGAGAACCAAAAAAATACCAAAAAAGAGAG 4549
230 pValArgLysSerValTyrPheProThrLeuArgHisGlu11eGluGluLeu 246
   :::::::::::::: ::::::::::::::::::::
4550 TAAGCTTATATTTAAGATAGAGAGACCAAAAGTGAATATAGTACAGATT 4599
247 .....11eLeuGluLysLeuLeuLysLe 254
   :
4600 GAATTCAGAGATGAGATATAGAGAGAGATGATACAAAGATATACAGA 4649
254 uAspValAsnAlaSerArgGlnGly11eGluAspAlaGluGluThrAlaT 271

```

```

4650 ACATATACAGACAAATATA.....GTTAGAGATATATAGATGAAATATAG 4694
271 1eGluThrCysGlyTyrThrAspSerThrGluGlyLeuPheAsnMetAsp 287
   :::::::::::::: ::::::::::::::::::::
4694 ATGAGAGATATA...GGTAAAGATCAAGATGAGATATATATATATAGCT 4740
288 GluAspGluGluThrGluHisGluThrLysAlaGlyProGluArgGluAs 404
   :::::::::::::: ::::::::::::::::::::
4741 CAAAAAGAGAAAGCGATTCAGAAAGATTAAGATCAAAAAAATAAATAA 4796
304 pGluMetValHisProValAlaGluAlaGluAsp11eLeuMetSerLeuLys 421
   :::::::::::::: ::::::::::::::::::::
4791 AAAAAAGATGAA.....GAAAGCTGTAGTGGCTTTAAAAAAATAG 4831
321 11eLeuSerTyrMetLysAspValLysTyrValAspGlyLysValAspAs 437
   :::::::::::::: ::::::::::::::::::::
4832 TAGACCAAGATATATGAAATATGTTAAAAATTCATTAAGAGATTCATAA 4884
338 GlyLysThrLysAspLeu.....TyrArgAspLeuLeuAsu11ePheAs 452
   :::::::::::::: ::::::::::::::::::::
4882 CAAGTATCTAAAGCTTCAGATATCAAAAAATGATGATTAATGTTTAA 4931
352 pLysLeuLeuLeuProThrHisAlaSerCysHisValGluPhePheMetP 469
   :::::::::::::: ::::::::::::::::::::
4932 ATAA.....AATTCAGATTTTIT..... 4950
365 1eTyrLeuLysSerPheLysGluGlyPheAlaAlaPheLeuGluHis 485
   :
4951 .....AGTAAAGCTTAAAAAG 4985
365 1eTyrLysLysLeuGluAspProSerAsnProAla11e1eArgGluAl 402
   :::::::::::::: ::::::::::::::::::::
4986 TTGCTAAAAAATATAAAGTATTTGCTGACATTCATATCTCCGCTGTC 5016
402 uAla.....GlyAsnTyr11eGlySerPheLeuAlaArgAlaLysPhe11eP 418
   :::::::::::::: ::::::::::::::::::::
5016 ACATATTCATCATATATATAGATTTGGCTCTTTCATTTTCTTATTTCA 5065
418 pGluLeu11eThrValLysSerCysLeuAspLeuValAsnThrPheHis 434
   :::::::::::::: ::::::::::::::::::::
5066 CATGTGTAAACAAAGCTTCTTCAAGTACTTATATTCACAAAGATGACAA 5116
435 11eTyrLeuAsnAsnGluAspSerGlyThrLysAlaPheCysAspValAl 451
   :::::::::::::: ::::::::::::::::::::
5116 ACTATCAAAHAAAAAATAAGAGAGCA..... 5149
451 1eLeuHisGlyProPheTyrSerAlaCysGluAlaValPheTyrThrPheV 468
   :::::::::::::: ::::::::::::::::::::
5149 .....CGTTTATTTA.....TTG 5155
468 uPheArgHisLysGluLeuLeuSerGlyAsnLeuLysGluIleLeu11e 484
   :::::::::::::: ::::::::::::::::::::
5156 TATTGAT.....ATTTTAAAGATTTAA.....CAT 5184
485 TyrLeuGlnSerLeuAsnPheGluAlaGluValMetSerGluLeuAsnTr 501
   :::::::::::::: ::::::::::::::::::::
5185 TATTTCACAAATATTAAGAGAAATTTAGTAAAGAAAAAATAATATATG 5234
501 pLeuLys11eCysLeuProSerValValAsnPhePheAlaVal11eThrA 518
   :::::::::::::: ::::::::::::::::::::
5235 AATGAGAGTAAACAAACAAAGCTGAGAAAAAGAGTATGTAAATGATCA 5284
516 1eLysTyrGluGluValPheCysLysTyrThr11eGluThrGluAsnMet 534
   :::::::::::::: ::::::::::::::::::::
5285 ATAAAAAGAGAGAGAA.....ACTAAAGTTGATAAAAATATATATAA 5325
535 GluMetLeuProVal11eArgSerThr 543
   :
5326 .....GTACCGAAAAAAAGAGAGAGAG 5346
seq_name: /cgp2_6/prodata/1/uu/6b_60m.seq:US-08-973-462-1

```



```

5484 .....CGTTTATTA.....TTG 5399
468 aIhPqRtHsIySgInIeUeUSeRgIyASnIeUySgIuGlyLeUgIn 484
5400 TATTATAT .....ATCTTAAGATTATAA.....CAT 5428
485 TyLeUgInSeRLeUaSnPheGluArqLeValMeISeRgIuLeUaSnPr 501
5429 IATTACAACAATATGAAACAAAATTTAGTAAGCAAAAAATATATCT 5478
501 oIuUySgIeCyLeUeProSeRValIaUaSnPheGluAlaIeIeThA 518
5479 AATAGAGATACAAACAAAGCTGACAAAAAGGTAATGTCAGAGTAAACA 5528
518 sOlySTyRgInIeUaIhPheGlyTyRtIleGluArqASnASnArq 534
5529 ATAAACCCAGAAACA.....ACTAAAGTTGATAAAATATAA 5569
535 GlMeIeUeProValIeArqSeRtHr 543
5570 .....GTACCCAAAAAAGAAAGACG 5590

seq_name: /c9n2_6/prodata/1/nu/6A_Comb.seq:US-08-776-265-4

seq_documentation_block:
: Sequence 4, Application US/08776265
: Patient No. 6001631
: GENERAL INFORMATION:
: APPLICANT: BLANCHET, Francis
: APPLICANT: CAMERON, Beatrice
: APPLICANT: CROZET, Joel
: APPLICANT: FAMECHON, Alain
: APPLICANT: FERREIRO, Lucia
: TITLE OF INVENTION: No. 6001631el Topoisomerase IV, Corresponding
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: ADDRESS: Dunney, L.L.P.
: STREET: 1300 I. Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/776,265
: FILING DATE: 24-JAN-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Finnegan, Carol P.
: REGISTRATION NUMBER: 42,220
: REFERENCE/DOCKET NUMBER: 03806, 0394-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4444
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2402 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-776-265-4

alignment_scores:
Quality: 98.50 Length: 470

```

```

Percent Similarity: 45.339 Percent Identity: 17.660
Ratio: 0.462
alignment_block:
US-09-932-678-2 x US-08-776-265-4
Align seq 1/1 to: US-08-776-265-4 from: 1 to: 2402
79 LeuAspProAsnIleUySAspAspInIleIAsnTrpLeuLeuIuh 95
111 .....
906 TCGAAAAAGATGACAGATGATCAATCAAAAATATATCT.....TA 949
95 eArqSeRSeIleMeTyRtIeUeTyRLeUySAspPheGluLeuIeSeR 112
.....
950 TAAAGCTCTGATTTACAGATTTCATATAATTC.....ACATGCTGCTA 996
112 IeIeUaRqLeUeProIhPheASnAISeRgInIhValIaIuGlu 128
11 .....
997 TATGATGATGTTGTCGAAATTTATGGGTATATGTAATATATATCT 1046
129 TyLeUaIaIhPheGlySAsnIeUaISeRAlaGInIhValIhPheUa 145
111111
1047 TATTGATCATCAATTCAGGTTGTCGAAATATGACGACATT..... 1091
145 qProCySLeUSeRMeIleUaISeRtHsIhPheValProIeArqValIle 162
1091 .....
162 IeIySgInIyASpValISeRAspSeRAspAspIuAspAspAsn 178
1061 .....
178 IeUeUaIaASnIhPheAspInIeUySAsnAlaIeUeUeIeIeAlaAr 195
1092 .....GAAATTAGATTAATGCTGAAAAACGATGATGATATGATGAG 1132
195 gTyRValProSeRtHrProIhPheUeUeProIeUaIyGluGlySP 212
113 .....TTTGCAATTTTACATAAAGTAA 1167
212 hArPheUaIArqLySSeRgInArqIhLeUgIuCyTyRValIhISn 228
1168 TCGAATTGATTCGACGCTCTAAACAAAGCGTACGCT.....AAAGAAAC 1214
229 IeUeUaRqIeSeRValIhPheProIhLeUaRqSeRtHsIleUeUe 245
1215 CIIATGAGATATACGACITTCAGAGAACACAGCGTCATACCAATGTAA 1264
245 oUeUeIleUySgIeUeUySLeUaSpValASnAlaSeRArqGlu 262
1265 GTACGCTTAATGCTTTAACCAACAGTCAATGATGATGCTGAGAGG 304
262 TyLeUaSpAlaIuGluIhArAlhPheIhPheGlyTyRValhAsp 278
1315 AACATAAGAACTTACAGCTTAATCAAAATATAGCTCATATTCAT 1364
279 SeRtHrGluGlyePheASnMeTAspIuAspIuIuhTrpIuhSeR 295
1365 AACATGATGATAIATGATGATTAATAAAGAAATATGAAAGAAATTA 1414
295 oThRySAsnIaGlyProGluArqLeUaSgIeUeValIhISpValIa 412
1415 AAGGAATTCAAATCTGAACGACGCTTTAATGAAGTAGAATTCAG 1464
312 Ty...ArqLeUaSP.....IleUeUe 418
1465 AATATAAAATGAAAGAAAGCTATAGTGGCTAGTACGAAAGATATTA 1514
318 SeRtHrVal..... 421
1515 AATATGACGCTATGATATATTAAGATGACTGCTATGCTGAGCTTAA 1564
422 LeuSeTyRMeIeUaSpValCySTyRValAspVal.....L 444

```













```

476 ATTCAG...GTAATAGAGAAATATCTGATTCAGGTAGTAAATATGTC 522
138 efaIagInThrValPheLeuArgProCysLeuSerMetIleAlaSerHis 154
523 ATGCGCTGGCAGGTCTTCCTCCGCCGGTGCATTCGATTCCTGTTTGTAAT 572
155 .....PheValProProArgValIle..... 161
573 TGTCTTTTAACAGCGATCGGATTTGCTGTGGCTACGCCGCAATCAGC 622
162 .IleLysGluGlyAspValAspValSerAspSerAspAspGluAspAsp 178
623 AATGAAATAACCGTTTGCTTCATCGCAGCATTTTTCATGACGACGCTAATG 672
178 sOleuPro 180
673 GCTGACCT 680

```

seq\_name: /cgn2\_6/prodata/1/lna/6R\_COMB.seq:US-09-082-649B-81

seq\_documentation\_block:

```

: Sequence 81, Application US/09082649B
: Patient No. 6339068
: GENERAL INFORMATION:
: APPLICANT: DAVIS, Heather L.
: APPLICANT: KRIEG, Arthur M.
: APPLICANT: SCHOF, Joachim
: APPLICANT: WU, Tong
: TITLE OF INVENTION: Vectors and Methods for Immunization or
: FILE REFERENCE: 01039/7909
: CURRENT APPLICATION NUMBER: US/09/082,649B
: CURRENT FILING DATE: 1998-05-20
: PRIOR APPLICATION NUMBER: US 60/047,213
: PRIOR FILING DATE: 1997-05-20
: PRIOR APPLICATION NUMBER: US 60/047,209
: PRIOR FILING DATE: 1997-05-20
: NUMBER OF SEQ ID NOS: 85
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 81
: LENGTH: 1360
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: plasmid DNA mutant Kanamycin resistance gene
US-09-082-649B-81

```

alignment\_scores:

```

Quality: 96.50 Length: 136
Ratio: 1.322 Gaps: 6
Percent Similarity: 53.676 Percent Identity: 28.676

```

alignment\_block:

US-09-932-678-2 x US-09-082-649B-81 ..

Align seq 1/1 to: US-09-082-649B-81 from: 1 to: 1360

```

58 GIValleuLeuLysTyrLysGlyGluThrAsnAspPheGluLeuLeu 74
|||||
309 GAGTTGTTTGAACATGGCAACGTAACGTTGCCAAT..... 447
74 uLysAsnGluLeuLeuAspProAspIleLysAspAspGluIle...IleA 90
|||||
348 .....GATGTTACAGATGAGATGGTCAGACTAA 375
90 snIlePleuGluGluPheArg.....SerSerIleMetLysLeuThrLys 104
|||||
376 ACTGCGCTACAAATATTATGCTCTTCGCCACATCAACGATTTTATACCT 425
|||||
105 AspPheGluGluGluSerIleIleLeuArgLeuProThrPleuAsnArg 121
|||||

```

```

426 ACTCTTATGATGCAAGTTTATTCACCAATCGGATCTTTGAAAAATAT 475
121 sSerGluThrValValGluLysIleValAlaPheLeuThrValAsnLeuValS 148
476 ATTCAG...GTAATAGAAATAATCTGATTCAGGTAAATAATATGTC 522
138 efaIagInThrValPheLeuArgProCysLeuSerMetIleAlaSerHis 154
523 ATGCGCTGGCAGGTCTTCCTCCGCCGGTGCATTCGATTCCTGTTTGTAAT 572
155 .....PheValProProArgValIle..... 161
573 TGTCTTTTAACAGCGATCGGATTTGCTGTGGCTACGCCGCAATCAGC 622
162 .IleLysGluGlyAspValAspValSerAspSerAspAspGluAspAsp 178
623 AATGAAATAACCGTTTGCTTCATCGCAGCATTTTTCATGACGACGCTAATG 672
178 sOleuPro 180
673 GCTGACCT 680

```

• •

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 30, 2002, 02:33:51 : Search time 2612.22 seconds

(without alignments)  
16566.781 Million cell updates/sec

Title: US-09-932-678-1

Perfect score: 2068

Sequence: 1 aagagagctgctgctggaag.....tgacatttgatccccc 2068

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595412

Minimum hit seq length: 0

Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb\_bat: 2: gb\_hgt: 3: gb\_hgt: 4: gb\_hgt: 5: gb\_hgt: 6: gb\_hgt: 7: gb\_hgt: 8: gb\_hgt: 9: gb\_hgt: 10: gb\_hgt: 11: gb\_hgt: 12: gb\_hgt: 13: gb\_hgt: 14: gb\_hgt: 15: gb\_hgt: 16: gb\_hgt: 17: gb\_hgt: 18: gb\_hgt: 19: gb\_hgt: 20: gb\_hgt: 21: gb\_hgt: 22: gb\_hgt: 23: gb\_hgt: 24: gb\_hgt: 25: gb\_hgt: 26: gb\_hgt: 27: gb\_hgt: 28: gb\_hgt: 29: gb\_hgt: 30: gb\_hgt: 31: gb\_hgt: 32: gb\_hgt: 33: gb\_hgt: 34: gb\_hgt:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No. Score Match Length DB ID Description

1	2068	100.0	2068	9	AF227156	AF227156	Human sapi
2	2020	97.7	3756	9	HS62272056	AX046047	Human sapi
3	2018.4	97.6	2040	6	AX046047	AX046047	Human sapi
4	1711.6	58.6	135873	2	AF012547	AF012547	Human sapi
5	1211.6	58.6	192815	9	AC013444	AC013444	Human sapi
6	567	27.4	2493	9	AK055742	AK055742	Human sapi
7	506.4	24.5	1177	9	BC009198	BC009198	Human sapi
8	420.8	20.3	1461	6	AX274907	AX274907	Human sapi
9	420.8	20.3	2627	9	BC006441	BC006441	Human sapi
10	208	10.1	129473	2	AC004018	AC004018	Human sapi
11	208	10.1	143900	2	AC092137	AC092137	Human sapi
12	192	9.3	175691	2	AC007615	AC007615	Human sapi
13	192	9.3	191496	2	AC009130	AC009130	Human sapi
14	192	9.3	208008	2	AC009093	AC009093	Human sapi
15	190.8	9.2	187643	2	AC092852	AC092852	Human sapi
16	188.8	9.1	173166	2	AC092852	AC092852	Human sapi
17	188.8	9.1	179150	2	AC017077	AC017077	Human sapi
18	188.8	9.1	202004	9	H04F001549	H04F001549	Human sapi
19	187.2	9.1	170611	9	AC106788	AC106788	Human sapi
20	158	7.6	143900	2	AC092137	AC092137	Human sapi
21	153.2	7.4	191496	2	AC009130	AC009130	Human sapi
22	153.2	7.4	196413	2	AC097269	AC097269	Human sapi
23	153.2	7.4	247331	2	AC097268	AC097268	Human sapi
24	146.4	7.1	2112	3	AY061123	AY061123	Human sapi
25	140.6	6.8	179150	2	AC017077	AC017077	Human sapi
26	131	6.3	293	9	HS151834	HS151834	Human sapi
27	127	6.1	78347	2	AC022642	AC022642	Human sapi
28	125.4	6.1	53814	2	AC107769	AC107769	Human sapi
29	118.6	5.7	824	2	AC107769	AC107769	Human sapi
30	115	5.6	53814	2	HS343405	HS343405	Human sapi
31	108.6	5.3	821	9	HS343405	HS343405	Human sapi
32	91.4	4.4	108040	2	AC068150	AC068150	Human sapi
33	91.4	4.4	171940	2	AC092119	AC092119	Human sapi
34	91.4	4.4	173166	2	AC092137	AC092137	Human sapi
35	91.4	4.4	316296	2	AC092285	AC092285	Human sapi
36	70.8	3.4	78347	2	AC022642	AC022642	Human sapi
37	62.8	3.0	405	11	G27379	G27379	Human sapi
38	62.8	3.0	6287	2	AC017809	AC017809	Human sapi
39	62.8	3.0	96018	3	AE004783	AE004783	Human sapi
40	62.8	3.0	172748	3	AC018484	AC018484	Human sapi
41	62.8	3.0	176424	3	AC011757	AC011757	Human sapi
42	59.6	2.9	174531	2	AC023752	AC023752	Human sapi
43	51.4	2.5	125020	9	AF429415	AF429415	Human sapi
44	48.2	2.3	8797	6	AF027319	AF027319	Human sapi
45	48.2	2.3	10338	4	BC054856	BC054856	Human sapi

## ALIGNMENTS

RESULT 1  
AF227156  
LOCUS Homo sapiens RRM3 mRNA, complete cds.  
DEFINITION AF227156  
ACCESSION AF227156  
VERSION AF227156.1 GI:7670099  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
AUTHORS Moorefield B., Greene E.A. and Reedor R.H.  
TITLE RNA polymerase I transcription factor RRM3 is functionally  
conserved between yeast and human  
PROC. Natl. Acad. Sci. U.S.A. 97 (9), 4724-4729 (2000)

JOURNAL MEDLINE  
20243763  
2 (bases 1 to 2068)

REFERENCE  
AUTHORS Moorefield B., Greene E.A. and Reedor R.H.  
TITLE Direct Submission  
JOURNAL Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, USA  
Location/Qualifiers



Db	1461	CAGCATATGACATCCGGCTGGAGGAGACTCAGTGCACATGTGCACAAACCCTTGACA	1740
OY	1741	CGTCCTCCGCCCTTTGATCTCCCTGTGTGTGTGAAGAGTCAAGAATLTCATGATCTATT	1800
Db	1741	CGTCCTCCGCCCTTTGATCTCCCTGTGTGTGTGAAGAGTCAAGAATLTCATGATCTATT	1800
OY	1801	atcagatatggaagaacataagtgtctgaagaqctacacgaqgtltaaaagaccatgaada	1860
Db	1801	ATCAGGATATGGAGAACATAGACTGCTGAAGAGCTACAGSAGTTCACAAACCCATGAAA	1860
OY	1861	aaggacaatggaaaattgaagatgatgaacttcgaaaggguaatgccccagaatga	1920
Db	1861	AAGGACATAGTAAGAAATGAAATGATGACTTTCTGAAAGGGGAAGTCCCCAAGATGATA	1920
OY	1921	cagtgatattgagatcacacacaaagctccctttgagacagcatcttccaagtlctcaagtag	1980
Db	1921	CCGTGATATGGAGATCACACCAACCTCTTTGACAGACGATTTCCGAAGCTCTTCAAGTAGTG	1980
OY	1981	tggagctcccaccccgatgtgatcatgagacccagatccgcctctagagaaaaatttga	2040
Db	1981	TGGGCTCCGCAACCGGCTGTGTACATGCAACCCAGCTCCCTCTACGGCGAATAATTGTGA	2040
OY	2041	ctgagatgtaacatttgqaattccccat	2068
Db	2041	CTGAGATGTACATTGGGATTCCCAAT	2068
RESULT 2			
HSA272050			
LOCUS	HSA272050	3756 bp	mRNA linear PRI 07-SEP-20000
DEFINITION	Homo sapiens mRNA for transcription initiation factor 1A protein (TIF-1A gene).		
ACCESSION	AJ272050		
VERSION	AJ272050.1	GI:10046713	
KEYWORDS	TIF-1A; transcription initiation factor.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	Bodem,J., Hoffmann-Rohrer,U., Koss,W., Delius,H., Vingron,M. and Grunert,I.		
TITLE	Cloning and functional characterization of transcription initiation factor TIF-1A, a growth-dependent regulator of ribosomal RNA synthesis		
JOURNAL REFERENCE	Unpublished		
AUTHORS TITLE	2 (bases 1 to 3756)		
JOURNAL	Bodem,J.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle		
TITLE	II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg, GERMANY		
FEATURES			
SOURCE	Location/Qualifiers		
gene	1..3756		
CDS	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	23..3756		
	/gene="TIF-1A"		
	23..1978		
	/gene="TIF-1A"		
	/function="transcription initiation factor"		
	/codon_start=1		
	/evidence="experimental"		
	/product="transcription initiation factor 1A protein"		
	/protein_id="CA007955.1"		
	/db_xref="gi:10046714"		
	/translation="MAAPLLHTRLPGHAASSAVKLGASRTGISNNRALENDFFNS		
	PERKTRFGQATEVEALLKKKGTEINDSEILKNLDLPDKDQIIMNLEFRSSIIMYLIEPSSIMTL		
	KDFPOLLSITLRPLMLRSOTVVEEELASGVASQIVLRPTSLMSHVPWVPPV		
	LIKEDVASISDDDEDLDPANDICHERALQIIARVSPFWFLMDITIVKPPVRSR		
	RKLKCYVAHNILIRLSVYPTLRHLIELLIERLIKDVNASOGIEDAEAFATQCGS		
	TOSTELLRNMEDETFETFKAGPERIDOMVHVAREDTLMSLVSTPKDVCYDGCK		
	VNGKTKDLIRKQLINFDKLLPHASCHVQFMFVLVSFKIGFAEALDELHKRKLOD		

[illegible]





















```

consensus quality: 16040 bases at least Q30
consensus quality: 14741 bases at least Q20
Estimated insert size: 21100; adapter-trim estimation
Estimated insert size: 142200; sum-of-reads estimation
Quality coverage: 4.0 in Q20 bases; adapter-trim estimation
Quality coverage: 5.94 in Q20 bases; sum-of-reads estimation
NOTE: This is a "working draft" sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
    *
    *      1      1034: contig of 1033 bp in length
    *      1034      1134: gap of unknown length
    *      1134      2547: contig of 1414 bp in length
    *      2548      2647: gap of unknown length
    *      2648      4205: contig of 1558 bp in length
    *      4206      4206: gap of unknown length
    *      4306      6564: contig of 2259 bp in length
    *      6565      6665: gap of unknown length
    *      6665      9065: contig of 2401 bp in length
    *      9066      9166: gap of unknown length
    *      9166      12059: contig of 2894 bp in length
    *      12060      12160: gap of unknown length
    *      12160      19295: contig of 7136 bp in length
    *      19296      19396: gap of unknown length
    *      19396      25668: contig of 6273 bp in length
    *      25669      25669: gap of unknown length
    *      25669      31013: contig of 5245 bp in length
    *      31014      31014: gap of unknown length
    *      31014      36555: contig of 5442 bp in length
    *      36556      36556: gap of unknown length
    *      36556      45338: contig of 8583 bp in length
    *      45339      45439: gap of unknown length
    *      45439      52016: contig of 7480 bp in length
    *      52017      53016: gap of unknown length
    *      53016      53019: contig of 12702 bp in length
    *      53019      65720: gap of unknown length
    *      65721      65821: gap of unknown length
    *      65821      76110: contig of 10289 bp in length
    *      76110      76210: gap of unknown length
    *      76210      85449: contig of 9230 bp in length
    *      85450      85449: gap of unknown length
    *      85449      101620: contig of 16081 bp in length
    *      101621      101720: gap of unknown length
    *      101721      122163: contig of 20443 bp in length
    *      122164      122164: gap of unknown length
    *      122164      143900: contig of 21637 bp in length
    *
FEATURES             location/Qualifiers
Source               1..143390
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="16"
                    /clone="RP11-144H15"
                    /clone_11b="RPCT human BAC library 11"
BASE COUNT          37791 a 34128 c 34320 g 35886 t 1775 others
ORIGIN
Query Match         10.1%; Score 208; DB Z: Length 143900;
Host Journal Similarity 100.0%; Prod. NO. 5.7e-46;
Matches: 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14f : agagcctatgaaagatgaatatgtacttcttaaaaggagaagtgcgcccaaatgafA 1920
|||||
Db 2733R AGGCAATAGTGGAGATGAATGTACTTCCTCGAAAGGCGAAGTCCTCCAGAACATATA 27279
|||||
QY 14d : ccgggatgagatccacccaagtctcttgacaagcatcttcgaagtctctgaatgagC 1980
|||||
Db 2727 d CCGTGATGGAGATGACAGCAAGCTGCTTTGACAGCATTTCTCCAGAGAGCTTCAAGTAAG 27219
|||||
QY 14e : taagattcccaccataatctatatacatatgaaccccaatccccctcttttgtatgaaaatcttAgA 2040

```

















```

634 roSetSerValGlySerProValLeuTyrMetClnProSerPro 650
|||||
1969 CTCACAGTAGTGTGGCTGGCCACCTGTGTGACATGACACCACTCC 2018
651 Lcu 651
|||
seq_name: gb-pr:HS272050
seq_documentation_block:
locus:
DEFINITION Homo sapiens mRNA for transcription initiation factor 1A protein
(TIF-1A gene).
ACCESSION AJ272050
VERSION AJ272050.1 GI:10046713
KEYWORDS TIF-1A; transcription initiation factor.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Gracilata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Bodem, J., Hoffmann-Kohrer, U., Koss, W., Dellius, H., Vingron, M. and
Gromm, I.
TITLE Cloning and functional characterization of transcription initiation
factor TIF-1A, a growth-dependent regulator of ribosomal RNA
synthesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3756)
AUTHORS Bodem, J.J.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2000) Bodem J.J., Molekulare Biologie der Zelle
II, Deutsches Krebsforschungszentrum, INF.280, 69120 Heidelberg,
GERMANY
FEATURES
source
1..3756
location/qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
23..3756
/gene="TIF-1A"
23..1978
/gene="TIF-1A"
/function="transcription initiation factor"
/codon_start=1
/evidence-experimental
/product="transcription initiation factor 1A protein"
/db_xref="GI:10046713"
/translation="MAAPLHPRLPGPAASSAVKKLGRIGISMRALENDFPNS
PRAKTVRGGVTEVLKLYKKGENDFELKNGLVEDDIDDDIIMWLEFSSIMYL
TKDFEOLISITILRLPLNRSQLVVEEYLAEGINVSQIVFELPPQSLMSIEFVPRV
IIEGVIVVSDSDDDNLPAHPFCHALOLIARYVSPFPLMPILVEKPEPVRS
ERTLECVNHLIRISVPELIRHELELIPKLIKLVNASRQGLDAEETATQCGG
TSTGELFEMODETEHETKASPRKIDQWHPAEHDLIMLIVISYKDDCYVNGK
VNGKIKOLYRLINIEPKLIKPHASCHVPEHYICSKRIGEAATLEHMKRIOD
PSNRPAIRQAGNITGSLAKAKTPLLTKSCDILVNLHITYLNQDSKRAEDV
ALDGPYSAGVAPYXEVFRKQLISGLKEGLVDLSNEFIVMSQLNPKILPS
VNEFEALITNKQLVFCYTIIEKNNRQMLPEVIRSTAGDSVOICTNPDLTFPEPCV
LKSKKFIIPYVQWEDMSAELOEFKPKMKDIVEDEDDDFIKGVPNDIVIGTTP
SSHDHFRSPSSSSVSPVLYMPSPL"
3701R
/gene="TIF-1A"
/evidence-experimental
BASE COUNT 1060 a 692 c 844 g 1156 t 4 others
ORIGIN
alignment_scores:
Quality: 3398.00 Length: 651
Ratio: 5.220 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.846

```

```

alignment_block:
US-09-932-678 2 x HSA272050
Align seq 1/1 to: HSA272050 from: 1 to: 3756
1 MetAlaAlaProLeuLeuHisThrArgClnProGlyAspAlaAlaAla 17
|||||
23 ATGATGGGAGTGGTCTCTTCACACCGCTTCCCGGAGATGGCGCTTC 72
|||||
17 ProSerAlaValIleLysLysLeuAlaSerArgThrGlyLeuSerAsn 44
|||||
72 GTCCTTCGACATTAAGAAACCTGGGCGCTGACAGATGGCATTCAGATA 122
|||||
44 CTATGAlaLeuGluHisAspPheAsnSerProArgLysIleVal 50
|||||
122 TGGTGCATTACAGATGACTTTTCAATTCCTCCCAAGAAAGAAATGAT 172
|||||
51 ArgPheGlyGlyThrValThrGluValLeuLeuLysTyrLysLysGly 67
|||||
173 GCGTTCGCGAATGATACAGCAAGCTTCCGACATGAAAAACGCTCA 222
|||||
67 GThrAsnAspPheGlyLeuLeuLysAsnGlnLeuLysAspProAsp 84
|||||
222 AACCAATATACCTTACAGTTGTGAAACCAACAGCTGTTAACATCAACA 272
|||||
84 YAspAspGlnIleIleAsnTrpLeuLeuLysArgSerSerIleMet 100
|||||
272 AGCATGACCAAGATCATCAACTGGCTGCTAGAAATTCCTCTTATATG 322
|||||
101 TTTGluThrLysAspPheGlyGlnIleSerIleGlnGluArgAsp 117
|||||
322 TACTTGACAAACAGCTTGGACCAACTTTCAGATATATTTACAGTTGCC 372
|||||
117 GTrpLeuAsnArgSerIleThrValValGlnGluTyrLeuAlaPheLeu 134
|||||
372 TGGTTCGATAGAAAGTCAACAGTAGAGCAAGATTGGCTTCTTCTG 422
|||||
134 LysAlaLeuValIleArgAlaGlnIleValPheLeuArgProLysLeu 150
|||||
422 GTAATCTGTATCAACACAGCTGTTCCTTGAAGAGTGTTCAGATG 472
|||||
151 MetAlaSerHisPheValProProArgValIleIleLysGlnLysPhe 167
|||||
472 ALTGCTTCCCATTTGGGCTCCCGCCACATCATGATTAAGCAAGACGAT 522
|||||
167 LAspValSerAspSerAspAspGlnAspAspAsnLeuProAlaAspPhe 184
|||||
52 AATGCTTCAGATTCGATGATCAACATGCAATTCCTTGCAAAATTTTG 572
|||||
184 SThrLysHisArgAlaAlaGlnIleIleAlaArgTyrValProSerThr 200
|||||
572 ATACATGTCACAAACCTTGCAAAATATATAGATATGTGATATGATGACA 622
|||||
200 ProTrpPheLeuMetProIleGluValGlnLysPheProPheValArg 217
|||||
622 CGTGCTTTCATACCCATACAGTACGGTGGAAGAAATTCGATGTTGTC 672
|||||
217 sSerGluArgThrLeuGluLysTyrValHisAsnLeuGluArgIleSer 244
|||||
672 ACGAGCAAGACACAGATGATGATGATGATGATGATGATGATGATGAT 722
|||||
244 gThrPheProThrIleArgHisGlnGlnIleGlnIleGlnIleLys 250
|||||
722 TATATTTTCCAACTTCAAGCATCAAAATTTTGAGACTTATTTTCAAAAA 772
|||||
251 LLeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGlnAspAla 267
|||||
772 CTACTCAGCTTCAGATGATGATGATGATGATGATGATGATGATGATGAT 822
|||||
267 gThrThrAlaThrGlnThrLysGlyTyrAspSerThrGlnGlyLys 284
|||||
822 ATAAACACCACTTAAATTTGGTGGGACATGATTCATGAAAGATTTGT 872

```

[illegible]



|||||  
40 AIGGGGACCGCTTCACACCGCTTGGCGGAGATGGCCCTTC 89  
17 rSerSerAlaValLysLysLeuGlyValAspArgThrGlyIleSerAsn 14  
90 GTCTCTGCAGCTTAAGAAAGCTGGCGCGTCGGAGACTGGGATTCACATA 149  
44 cTAlaGAlaLeuGluAspAspPhePheAspSerProPheArgLysThrVal 50  
140 TGGTGGATAGAGAAATGATTTTCATTTCCGCCCAAGAAACCTCTT 189  
51 ArpPheGlyGlyThrValThrGluValLeuLeuLysLysGlyGly 67  
190 CGGTTGGTGAACTGGACCAAGCTTCCTGCAAGTACAAAAAGCGTGA 249  
67 uThrAspPheGlyLeuLeuLysAsnGlnLeuLeuAspProAspIle 84  
240 AACCAATGACCTTGAGCTGTGAGAACCCAGCTGTACATCCACACATA 289  
84 yAspAspGlnIleLeuSerPheLeuGluPheArgSerSerIleMet 100  
290 AGGATGACCAATCATTAATGCTGCTAGAAATTCCTTCATCATG 339  
101 TyrLeuThrLysAspPheGlyGlnLeuIleSerIleIleLeuArgLeu 117  
340 TACTTGACAAACACCTTGACCAACTTATCATATTATTAAGATTGCC 389  
117 oThrLeuAsnArgSerGlnThrValValGlnLeuThrLeuAlaPheLeu 134  
390 TTGCTTGATAGAAAGTCAACAGTACTGCAAGCATATTGGCTTTCTG 439  
134 LysLeuValSerAlaGlnThrValPheLeuArgProCysLeuSerMet 150  
440 GTAACTGTGATACCAACACACTGTTCTCCAGAACGTTGTCCACATG 489  
151 LLeuAspHisPheValProPheArgValIleIleLysGlnGlyAspVal 167  
490 ATTCCTTCCTCATTTGTGCTCCGAGATCATTTAAGCAAGCGCATG 539  
167 LAspValSerAspSerAspAspGluAspAspAsnLeuProAlaAsnPhe 184  
540 ACAATGTTCACTTGATGATGATGATGATGATGATGATGATGATGAT 589  
184 sPheCysHisArgAlaLeuGlnIleIleAlaArgTyrValProSerThr 200  
590 ACACATTCACAAAGCTTGCATAATATAGCAAGATATACACATCGACA 639  
201 ProThrPheLeuMetProIleLeuValGlnLysPheProPheValArgLys 217  
640 CCGTGGTTTCAGATGCAATACTGCTGCAAAAAATTTCCATTTCGAAA 689  
217 sSerGlnArgThrLeuGlnLysTyrValHisAsnLeuLeuArgLysSer 234  
690 ATCAGAGAGAACTGCAATGTTACGTTCACTTACTACTAGCATTAAGT 739  
234 aTyrPheProThrLeuArgHisGlnIleLeuGlnIleIleLeuLys 250  
740 TATATTTTCCAACTTGAGCAATGAAATTTCTGAGCTTATTTATGAAA 789  
251 LeuLeuLysLeuAspValAsnAlaSerArgGlnGlyIleGluAspAlaGly 267  
790 CTACCAAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 839  
267 uGluThrAlaThrGlnThrCysGlyGlyThrAspSerThrGlnGlyLeu 284  
840 AGAAACACCACTCAAACTTGTGTGTGAGACATTTCCACCAAGCATCT 889  
284 hAsnMetAspGluAspGlnGlnThrGlnHisGlnThrLysAlaGlyPro 300  
890 TTAATATGATGAAGATGAAGAAACTGAAATGAAACAAAGCTGCTCT 939  
301 GlnArgLeuAspGlnMetValHisProValAlaGlnArgLeuAspIleLeu 417  
|||||  
940 CAACCGCTCGACCAATGATGTGATGCTTACCTGAGGCGCTCGACATCT 989  
417 oMetSerLeuValLeuSerTyrMetLysAspValCysTyrValAspIle 344  
990 GATCTCTGCTTTTGTCTACATGAGGATATCTGCTATGATGATGATGAT 1049  
344 yValAspAsnGlyLysThrLysAspLeuTyrAlaAspLeuIleAsnIle 460  
1040 AGGTTGATACCGCAAAACCAAGATCTATATCGAGCTCATTAATAAT 1089  
351 PheAspLysLeuLeuLeuProThrHisAlaSerValHisValGluPhe 367  
1090 LTRACCAACTCTCTTTCCTACCACTGCTGCTGCTGCTGCTGCTGCT 1149  
367 eMetPheTyrLeuGlySerPheLysLeuThrAlaGlnAlaPheLeu 484  
1140 CAGTTTATACCTGAGTTTCAATTTGATGATTCGCAAGCATTTTTC 1189  
384 LHisLeuThrLysLysLeuGluAspProSerAspProAlaIleLeuArg 400  
1190 AACATCTCTGAAAAATTCGAAATCCAAATCAATCTGCTGCTGCTGCT 1249  
401 GlnAlaAlaThrValSerTyrIleGlySerPheLeuAlaValAlaLysPhe 417  
1240 CAGCTTCCTGAAATTAATTTGCAAGCTTTTGTGCAAGCATTAATTA 1289  
417 ePheLeuIleThrValLysSerCysLeuAspLeuValAsnThrPhe 434  
1290 TCTCTCTATATACGTAATAATGCTGATCTTTGCTTATCTGCTGCT 1349  
434 LLeuTyrLeuAsnAsnGluAspSerGlyThrLysAlaPheCysAspVal 450  
1340 ACATATACCTTAATACCAAGATTTGCAAGCAATTCGCAAGCTTT 1389  
451 ALeuGlnLysLeuProPheTyrSerAlaCysGlnAlaValPheTyrThr 467  
1390 GCTTCATGATGATGATTTTACGATGCTGCTGCTGCTGCTGCTGCT 1449  
467 eValPheArgHisLysGlnLeuLeuSerLysLeuLysGlnGlyLeu 484  
1440 TGTATTAGACCAAGCATTTTACCGCAAGCTGCAAGACCTTTGC 1489  
484 LLeuLeuGlnSerLeuAspProGluArgIleValMetSerGlnLeuAsn 500  
1490 AATATCTTCAGATCTGAAATTTGAGCGGATGATGATGATGATGAT 1549  
501 ProLeuLysLysCysLeuProSerValValAsnPheProAlaAlaIle 517  
1540 CCGCTGAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1589  
517 PAsnLysTyrGlnLeuValPheCysTyrThrIleIleGlnLysAsnAsn 544  
1590 AAATATATATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1649  
544 TArgIleMetLeuProValIleArgSerThrAlaGlyCysAspSerValGln 560  
1640 GCGAATCTGCGCAATATGAGGATAGCGGAGAGAGAGAGAGAGAG 1689  
561 LLeuCysThrAsnProLeuAspIlePhePheProPheAspProValIle 567  
1690 ATCTGCAAAACCCCGAGACCTCTTTCCTGTTGCTGCTGCTGCTGCT 1749  
567 uLysArgSerLysLysPheLysAspProIleGlyGlnValIleProLys 584  
1740 GAAAGAGTCAAAAGAAATCATTTGATCTTATTAACAGTGTGAGAGACA 1789  
584 eSerAlaGlnGlnLeuGlnLeuPheLysLysProMetLysLysAspIle 600  
1790 LLeuGlnTGAAGAGTTACAGAGTTTCAAGCAAAACCATGAAAAACATA 1849  
601 ValGluAspGluAspAspPheLeuLysGlyGlnValProGluAsnAs 617  
1840 CTGAGACATGAGATGATGCTTCTGAAAGCGCAAGTCTGCTGCAATCA 1889







```

repeat_region /rpl_family="(GA)n"
12758..12785
/rpl_family="(TG)n"
13033..13054
/rpl_family="(CAAAA)n"
13195..13699
/rpl_family="LI"
13543..13567
/rpl_family="AT-rich"
14142..14305
/rpl_family="MIR"
14500..14534
/rpl_family="AT-rich"
14617..14735
/rpl_family="MIR"
14736..14836
/rpl_family="MER2_type"
14837..15106
/rpl_family="Alu"
15107..15317
/rpl_family="MER2_type"
15318..15385
/rpl_family="MIR"
15578..16066
/rpl_family="ENV1"
16072..16112
/rpl_family="AT-rich"
18192..18222
/rpl_family="AT-rich"
18612..18699
/rpl_family="Alu"
19181..19405
/note="match to EST H89091 (NID:q453918) yq02f04.s1"
19879..21714
/note="similar to katus norvegicus EST Y16973
(NID:q4727035)"
19890..20284
/note="match to EST H89197 (NID:q454024) yq02f04.r1"
19915..20369
/note="similar to Homo sapiens EST H6535886
(NID:q13527431)"
19943..20378
/note="similar to Homo sapiens EST AA213789 (NID:q1812416)
zr9a11.s1"
19923..20369
/note="similar to Homo sapiens EST BC428105
(NID:q13334811)"
20091..20378
/note="similar to Homo sapiens EST A1742153 (NID:q5110441)
w49f01.s1"
20158..20664
/note="similar to Homo sapiens EST AV704279
(NID:q10720608)"
20243..21062
/note="similar to Mus musculus EST H6295610
(NID:q13057417)"
20412..21044
/note="similar to Homo sapiens EST BE744354
(NID:q10148346)"
20567..20666
/note="similar to Homo sapiens EST BG778634
(NID:q14048951)"
20592..20963
/note="similar to Homo sapiens EST AA481295 (NID:q2210847)
zv02c05.f1"
20703..21362
/note="similar to Homo sapiens EST BG502871
(NID:q13464488)"
20798..21056
/note="similar to Homo sapiens EST AV645508
(NID:q3870522)"
20862..21269
/note="similar to Homo sapiens EST AA191111 (NID:q1779805)

```

```

misc_feature 20930..21593
/note="similar to Homo sapiens EST H6204664
(NID:q1097270)"
misc_feature 20930..21038

alignment_scores:
  Quality: 2049.00 Length: 645
  Ratio: 3.776 Gaps: 17
  Percent Similarity: 81.721 Percent Identity: 69.612

alignment_block:
US-09-932-678-2 x AC013444 ..
Align seq 1/1 to: AC013444 from: 1 to: 192815

17 GYASPAIAIAIAISerSerAlaValIysIysLeuGlyAlaIserAl 28
|||||  |||||||  |||||||  |||||||  |||||||  |||
19844 GSNAAAGAAAGAGGTTCAGTTCGTGGTCAAAAGCTGGATATGCAAG 19893
|||||  |||||||  |||||||  |||||||  |||||||  |||
26 GTGCGYIIGSerAsnMetArgAlaIuouIuAsnAspPheAsnSerP 45
|||||  |||||||  |||||||  |||||||  |||||||  |||
19894 GACCGATTTTCAGATATGTTGATTCAGAAATGATTTTCAGATTCG 19943
|||||  |||||||  |||||||  |||||||  |||||||  |||
45 rGPrARdysThrValArgPheGlyValThrValThrGluValLeuLeu 61
|||||  |||||||  |||||||  |||||||  |||||||  |||
19944 CCCCAAGAAAATTTTCATCTTCATCTTCATCTTCATCTTCATCTTC 19993
|||||  |||||||  |||||||  |||||||  |||||||  |||
62 LysIyTLysIysGlyGluThrAspAspPheGluLeuLeuLysAsnIle 78
|||||  |||||||  |||||||  |||||||  |||||||  |||
19994 AAGGACAAATTCATAAATAATTACCTTGAGCTGTGTGACAAAGCAAT 20043
|||||  |||||||  |||||||  |||||||  |||||||  |||
78 uGuaAspProAspIleLysAspAspGluIleLeuAsnTrpLeuLeuLeu 95
|||||  |||||||  |||||||  |||||||  |||||||  |||
20044 GAGAAATTCACACATAAAGCAATCCACATCTTGACCTGCTGACAA 20093
|||||  |||||||  |||||||  |||||||  |||||||  |||
95 hGAcSerSerIleMetTyrIleThrLysAspPheGluLeuLeuLeuSer 111
|||||  |||||||  |||||||  |||||||  |||||||  |||
20094 TTTCCTCTGATACAGTACTTCGCAAAACAGCTTCAGACCTCTCCAT 20143
|||||  |||||||  |||||||  |||||||  |||||||  |||
112 ILeIleuAlaIuouProTTPLeuAsnArgSerGluThrValIysIle 128
|||||  |||||||  |||||||  |||||||  |||||||  |||
20144 ATTACATTGAAATGCTTGGTGTGATATGAAATTCAGAAAGTGGTGA 20193
|||||  |||||||  |||||||  |||||||  |||||||  |||
128 uTyTLyLeuAlaPheLeuGlyAsnLeuValSerAlaIuouThrValPro 145
|||||  |||||||  |||||||  |||||||  |||||||  |||
20194 GTATTGGCTTCCTGGGGTACTCTGTGACACAGAA...GCTTCTGTTA 20243
|||||  |||||||  |||||||  |||||||  |||||||  |||
145 rProCysLeuSerMetIleAlaSerHisPheVal...ProProArgVal 160
|||||  |||||||  |||||||  |||||||  |||||||  |||
20244 GTCGCAAGTCCTACATATATTCCTTCCCATTCGATGCTGCTTCCTTA 20293
|||||  |||||||  |||||||  |||||||  |||||||  |||
161 ILeIleLysGluGlyAspValAspValIserAspSerAspAspIuAsn 177
|||||  |||||||  |||||||  |||||||  |||||||  |||
20293 AICATTAAAGAAAGCATGATTAATCTTCGAAATTCGATGATGATGAT 20342
|||||  |||||||  |||||||  |||||||  |||||||  |||
177 pGnIouProAlaAsnIleAspThrCysHisArgAlaIuouIleIleVal 194
|||||  |||||||  |||||||  |||||||  |||||||  |||
20342 TATCTCTCTCTACATATTCGACACAGCTGATACAAATTAAGAAAG 20391
|||||  |||||||  |||||||  |||||||  |||||||  |||
194 IArgTyValProSerThrProTTPheLeuMetProIleLeuValIle 210
|||||  |||||||  |||||||  |||||||  |||||||  |||
20391 CAGCATATTCGATCTGACACTGTGCTGCTTCGCAATCTGCTGCA 20439
|||||  |||||||  |||||||  |||||||  |||||||  |||
211 LysPheProPheValArgLysSerGluArgThrLeuGluGlyIysVal 227
|||||  |||||||  |||||||  |||||||  |||||||  |||
20444 AATTTCGATTTGTT...CAATCAAGACAGCAATGCAATGATATCTA 20493
|||||  |||||||  |||||||  |||||||  |||||||  |||
227 SASnLeuLeuArgLysSerValTyrPheProThrLeuAlaPheSerIle 244
|||||  |||||||  |||||||  |||||||  |||||||  |||
20487 TACTTACTTAAGATTTACTATATCTTTCATCTTCAGCTTACCAATTC 20535
|||||  |||||||  |||||||  |||||||  |||||||  |||
244 rGluIleuLeuIleIleGluLysLeuLeuLysLeuAspValAsnAla 260

```





Michael Thornton, Miranda Usal, Natasha van den Bosch, Jill Vardy,  
George Yang, Scott Zuydam, Marco Bortol.

clone distribution: MC clone distribution information can be found  
through the J.M.A.G.E. Consortium/JMNC at: <http://image.jim.gov>  
Source: IMC, Plate: 18 row: 4 column: 24

This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

FEA10623 Location/Qualifiers

Source:

1: 2627

Organization: "Homo sapiens"

Accession: "U00006"

Accession: "M001469" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"

Accession: "U00006" IMAGE: 351791"







```

source      1. 2112
            /organism="Drosophila melanogaster"
            /strain="y: cu bw sp"
            /db_xref="taxon:7227"
            /map="40D2-40D3"
gene         1. 2112
            /gene="CG3278"
            /note="alignment with genomic scaffold AE003763 gene does
            not completely overlap longest ORF"
            /db_xref="FLYBASE:FBgn0032988"
            178. 2013
            /gene="CG3278"
            /note="longest ORF"
            /codon_start=1
            /product="FLYBASE:FBgn0032988"
            /protein_id="AAL2867.1"
            /db_xref="GI:16768904"
            /translation="MSPTSKDVSLKTFSGYKKNQANAAVINKRSTPKREGLA
            KSVRALEERNHIVKEPTFLREAELESDPVIIKDAKIVHNLPEPTVVEAL
            SLNKKRSSELEAVTEFCLEAVTNNRYLPIGVSLIVHILPGMDESDWAGCPSE
            HVNELKPIHKVLDILTAVPAMDIIIDAIKPEPKPYTAGYLYNVIIMLEY
            KPIPELVOLYLOKLLIDYNAPKDEILEMDEKVEADTLEPIDVCNENPKS
            EKSVHVGKTIQDYCLELYKEPKCKLADNSSEBRKIANRIFMLIITPEVILP
            SHNHVQELIPLATSLRSVTSFAFDISTAKVONPWSAVIRHAAVGMASFTLSRAK
            FLPLSTVTPYIKETRMAMTYIDSDAVKONCSIKAMVYVSQVAFYLLAPRAKDL
            TASKDLLEQSLQSLAMQHLNPLRYCLAPVATFVATVYOLATYCHTLENNAR
            RMATVYHEKRMDELTSEFPDPYVLRKRYLETINMYVQSDNIDTEVYGGSNKY
            GHSRKKQSEMLEDEFLIDKRPKNFDLSQSEFQKQHPGSSP"
BASE COUNT      649 a      405 c      414 g      644 t
ORIGIN

alignment_scores
quality:      783.00      Length:      659
Ratio:      1.967      Gaps:      18
Percent Similarity: 60.395      Percent Identity: 30.501

alignment_block:
US-09-932-678-2 x AY061123 ...

Align seq 1/1 to: AY061123 from: 1 to: 2112

19 SerAlaValAllystleuglyAlaSerArgThrglyIleSerAsnMetArg 45
   |||||
229 TCTGGCGTAAAGCAAAATCAAGCAAAATGCGCGCATCAACAAGTTGCG 278
   |||||
35 gAlaLeuGluAsnAspPheAsnSerProPArglysthrValArgP 52
   |||||
279 A.....TTTTCACACCCCAAGAGAG..... 300
   |||||
52 hestglyThrValThrcValLeuLeuLeuLeuLeuLeuLeuLeuLeu 68
   |||||
301 .....GGGCTGCCAATACATCACTGCGCTGCCCTA.....CAAGAA 336
   |||||
69 AsnAspPheGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 82
   |||||
337 GCAAAATTCATCTGGTCAAGAGGTTCACTACTCTCTCGGGAACAACA 386
   |||||
82 pIleLysAspAspGlnIleLeuAsnTrpLeuLeuGluIupheArgSerI 99
   |||||
387 ACTCAGCATGATGAGCTGTGATATATCAATAAAGATGCTAAAGCATG 436
   |||||
99 lometLysLeuThrlLysAspPheGluLeuLeuLeuLeuLeuLeuLeu 115
   |||||
437 TGCATTAATTTACTCTCGAATTTCTTAAGCGTCTGAGCGCTTACTGTC 486
   |||||
116 leupTrpIleuAsnArgSerGlnThrValValGluGluTyrIleValArg 132
   |||||
487 CTTAAGTGAAGAAAGAAAGAGTTTCAGACATATACGAGCTTATACGAAT 546
   |||||
132 cleuGlyAsnLeuValSerAlaGlnThrValPheLeuArgProGlySer 149
   |||||

```

```

537 TGTATCGAACIAATGCTGACCCATATACATATTATCAATTCAGCTCT 586
149 ermeLleAlaSerHisPheValProProArgValIlelleLysGluGly 165
   |||||
587 CTAACACTAATTTGCGACAGCATTCCT.....GGA 615
166 AspValAspValIserAsp.....SerAspAspGluAspAspAsnLeuPro 180
   |||||
616 GATATGATATCATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 665
180 gAlaAspPheAspThrCysHisArgAlaLeuGluIlelleAlaArgTyr 197
   |||||
666 AAATGAGTGAAGAAACGATCAATCAATCAATCAATCAATCAATCAAT 715
197 alProSerThrProTrpPheLeuMetProIleLeuValGluLysPheTrp 214
   |||||
716 TCTTAAGGCTTCGATATATTATCATAGATCAATCAATCAATCAAT 765
214 PheValArgLysSerGlnArgTrpIleGluLysTyrValHisAsnLeuLe 230
   |||||
766 TATTCAAAAGCCGACGACATGTCAGCGCGAGGTATCTGTATTAAGTCT 815
235 GATGlleSerValTyrPheProThrLeuArgHisGluIleleuGluLeu 247
   |||||
816 TTCGTATTAAGAAATCAAGCGTATTTGCAAGCACTTGTTCGAACTAG 865
247 lelleGluLysLeuLeuLysLeuAspValAsnAlaSerArgGlnLysIle 263
   |||||
866 TTTTACAAAATCTTTATATATGATGTTAATGTTAATGTTAATGTTAAT 915
264 GluSpatArgIleuIleThrAlaThrGluThrCysGlyThrAspSerThr 280
   |||||
916 GAA..... 948
280 gLuoGlyLeuPheAsnMetAspGluAspGluLeuThrGluHisGluThr 296
   |||||
948 .....TAAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 961
297 .....LysAlaGlyProGluArgLeuAsp 304
962 TATTTGAATTAATGATGATGATGATGATGATGATGATGATGATGAT 1011
305 GlnMetValHisProValAlaArgLeuAspIleleuMetSerLeuVal 321
   |||||
1012 TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1061
321 lleuSerTyrMetLysAspValCysTyrVal.....AspGlyLysValAla 346
   |||||
1062 TATTAAGTTTTCATGCGCAAAATGTCGCTTAATGAAGAAATAGTACG 1111
346 SerAsnGlyLysThrLysAsp.....LeuTyrAlaGAspLeuIleAsnIlePhe 351
   |||||
1112 AACACGAGAAAGATGCCAAGCGATGTTAAGCTTACTATTATTCTTC 1161
352 AspLysLeuLeuLeuLeuProThrHisAlaSerCysHisValGlnPhePheMet 368
   |||||
1162 GACGAGGCTGTTATTCGACGACCAACACGACATCAATGCAATTCAT 1211
368 lPheTyrLeuLysSerPheLysLeuGlyPheAlaGluAlaPheLeuGlu 385
   |||||
1212 ATTATGACGATGATTCGCGGTCGCGATATTCGCAAGCTTTCTTATAT 1261
385 lserTrpLysLysLeuGluAspProSerAsnProAlaIlelleCArgGln 401
   |||||
1262 CTTATGACGAAAGTCAAAATGCGAATGATATGCTGATTTGATGAT 1311
402 AlaAlaLysAsnTyrIleLysSerPheLeuAlaAlaAlaLysPheIleP 418
   |||||
1312 GTTGATGAGGAAAGATGAGTATTTATCTGCGGCGACGATTTTCTG 1361
418 GluPheIleThrValLysSerCysLeuAspLeuValAsnTrpLeuHis 445
   |||||
1362 GTTAGCAGCTTACCTTCTACCTAAAGAGTTCAGACAGCTGGGTTATA 1411

```







56	.....	GhVAlleuLeuLysTyrllysAlaAsp	71
821	CTTCTCAAAATTC	CCCTCAAGATCTGGATGAAGAATGATCAACCCAA	870
71	hacUleuLeuLysAsnGlnLeu.....	AspPaspHe	84
871	TGGATATATGCAATCAACAGTAGCGCTCCGACAGAAATCTGAAGA	920	
84	LysAspAspGlnIleLeuAsnTyrLeuGlnIuPhaArgSerTHeu	100	
921	AlAAATGATGAAGATCTCAATATCTTTGCATATCTTATCAATATAT	970	
101	CTTyrLeuThrLysAspPheGlu.....	GlnLeuIleSerTHeu	114
971	CACAGAGATACAAATCCGACGGAACATTCCTCAACAAATCAATAA	1020	
114	euArGLeuProTyrLeuAsnArgSerGlnThrValValGlnTyrLeu	130	
1021	ACTTTGAGAAATGGATGGAAATTCCTCCACACACTCTAGATAAATATA	1070	
131	AlAPheLeuGlyAsnLeuValSerAlaGlnThrValIheLeuArgProCy	147	
1071	TATTCAATCAAAATCTTTGCTGGATATACCCAAAGGAGGACAAAGCT	1120	
147	SenSerMetIleAlaSerHISpHeValProProArgValIleIleArg	164	
1121	TTTCATGATACATGATATCTGTTTATTT.....	1140	
164	1AclYAspValAspValSerAspSerAspAspGluAspAspAsnLeuPro	180	
1146	.....	TTACCA	1150
181	AlaAsnPhaSerThrCysHisArg...AlaGlnGlnIleLeuAlaArgTy	196	
1155	AlCAAAACGACAGTGTCATCATCATATATGCTTAAATATTTCTTAAG	1200	
196	TYAlProSerThrProTyrPheLeuMetProIleLeuValGluLysPhe	213	
1206	GATTCCTCCCTCGAATGGCATTCATACATATTTGCCAAATTTTC	1250	
213	rPheValArgLysSerGluArgTyrLeuGlyCysTyrValHisAsnLeu	229	
1256	CAATTAATAAATGATACTCGACGAAATATGTTAATCACTCCAAATTG	1300	
230	IeuArgIleSerValTyrThrProThrLeuArgHisGluIleuLeuIle	246	
1366	CTGAATATGACAGCTGCTCTGTAGCTAGATTCGAAATTTGGTCTT	1350	
246	uIleLeuLysLeuLeuLysLeuAspValAsnAlaSerArgGln....	261	
1356	ATGATACAAAATAATATTTCAATTAATGTCAGTTAAATAATGAATAG	1400	
267	.....	GlyIleGluAspAlaGlnGlnThrAlaThr	271
1405	ATGAATTAGATGATGATAGATGATGATGATGATGATGATGATGATG	1450	
272	GTrnTyrGlyGlyThrAspSerThrGlnGlyLeuPheAsnMetAspI	288	
1456	GAGCATGACGACATTTACATGATCATTTCCAGTA.....	GAGCA	1490
288	uAspGluGlu.....	ThrGluHisGlnIleLysAlaGlyProG	301
149	TSATATGTAATAATTCCTGGACAGATAATGAAAGCTTAAGCTGCGACAG	1540	
301	Lu.....	ArgCysAspIleMetValHisTyrVal	310
154	CTGAAGCTAGTCAAGCTAGTGAAGAATATGACATCATTAAGATAG	1590	
1563	GAGGCTACTAACAATATATATCTGGAGCTTGACGCGAGAGGAAATCAACAACT	1640	

[illegible][illegible]



```

46 ..ProAlaIysThrValArgPheCly.....GlyThrValThr... 57
      ::::: |||::: |||
559 AAAAAAAGAGAGGTCAGTTTCAGATAGTACTGGCTGTAACCTT 608
      ::::: |||
57 ..... 57
609 CACCCAGAGAGATAAAAGATGAAGTCTTTCCAGCCGCAATGATAGAG 658
8 .....GluValLeuLeuIysTyrIlyslYslGlyIuhTrAsp 71
      ::::: |||
659 GTTCGTGAATTCGCCATACATATGATGATAGATATGATACCCAGCA 708
71 heGluLeuLeuIysAsnGluLeuLeu.....AspProspile 83
      ::::: |||
709 TCGCATATATTCGCATCATCATACAGCTTCCAGCAGACATCTTAACA 758
84 IysAspAspGluIleIleAsnTyrPleuLeuGluPheArgSerSerIle 100
      ::::: |||
759 ATTATATGAAGAATCTCAATATCTTTTCATATCTTACTAGTAATAT 808
100 tTyLeuThrIysAspPheGlu.....GluLeuIleSerIleIle 114
      ::::: |||
809 CACACGGATACATATCTCCAGGAGACTTCTCTCATACATCATCATTA 858
114 euArgIleuProTyrPleuAsnArgSerGluThrValGluIuhTrLeu 130
      ::::: |||
859 ACTTGACAGACTGGTGGGAATTCCTCCACACACTCTAAGTAAATATA 908
131 AlapheluGlyAsnLeuValSerAlaGluThrValPheLeuArgProcy 147
      ::::: |||
909 TATTTATCAAAAATCTTTGCTCGAGATATCCAAATGGGCAAGACT 958
147 sleuSerMetIleAlaSerHisPheValProProArgValIleIleLysG 164
      ::::: |||
959 TTGGATGACTCTGATCTGTTTAT..... 986
164 IuGIAspValAspValSerAspSerAspArgIusAspAsnLeuPro 180
      ::::: |||
987 .....TTACCA 992
181 AlaAsnPheAspThrCysHisArg...AlaLeuGluIleIleAlaArgTy 196
      ::::: |||
993 ATCAACACAGACAGTCTGTCAATGATATGCTTAATATTTCTTAAGAT 1042
196 rValProSerThrProTyrPheLeuMetProIleValGluLysPhe 213
      ::::: |||
1043 GATTCCTCTCGATTCGATTCATACATATTTGGCCAAATTTTTCCT 1092
213 roPheValArgLysSerGluArgThrLeuGluLysTyrValHisAsnLeu 229
      ::::: |||
1093 CAATATAAATCATACTCTGGGAAAGTTCGTTAATACACTCAAAATTG 1142
230 LeuArgIleSerValTyrPheProIleLeuArgHisGluIleLeuGluLys 246
      ::::: |||
1143 CTGAATATGACAGGCTACTGCTGTGACCTAGCATTCGCAATTTGGTCTT 1192
246 uIleIleGluIysLeuLeuLysLeuAspValAsnAlaSerArgGlu.... 261
      ::::: |||
1193 ACTGATAGAAAAGATTTTTCGATTCGATTCGATTAACAAATGATTAG 1242
262 .....GlyIleGluAspAlaGluIuhTrAlaThr 271
      ::::: |||
1243 ATGATATGATGATGATCTACATGACGATGATCTAGAGACACTGTCTTA 1292
272 GluThrCysGlyGlyThrAspSerThrGluGlyLeuPheAsnMetAspS1 288
      ::::: |||
1293 GAGGAGTCAGACGATTTAGATGATGATCTCAGCA.....CACCA 1330
288 uAspGluLys.....ThrGluHisGluThrLysAlaGlyProG 301
      ::::: |||
1331 TCATCATATAAATTCGTGGACACACTATGACAGTTAAGTCTCGAGCAG 1380

```

```

301 .. 301
1381 CCGAGGTAGTCAAGTGAATGAAATATATGACATGATTGAGATAG 1430
302 .....ArgLeuAspGluMetValHisProGlu 310
      ::::: |||
1431 GACGCTACTGAGAGATAATAATGAGAGTTGACCGACCGAGATCAAGAGACT 1480
310 lAlaGluArgLeuAspIleLeuMetSerLeuValLeuSerTyrMetLysA 1427
      ::::: |||
1481 TCCGACCAAAATGACACTCTATCTTACTTCTTAAACGACGATGAGAG 1530
327 sValCysTyrValAspGlyLysValAspAsnGlyLysThrLysAspLeu 344
      ::::: |||
1531 ACAAAGTACGCGGAG...AGTTAAAGACGTGAAGCTAGATATC 1577
344 TyrArgAspLeuIleAsnIlePheAspLysLeuLeuLeuProThrHisAl 160
      ::::: |||
1579 TCAACACTCTAAACACATATTTAAACCGCATGCTACCCACCTACTA 1627
365 aserCysHisValGluPhePheMetPheTyrLeuLysSerPheLysLeuG 1677
      ::::: |||
1628 TACCAAGTGAATTCATATCATATTCATGATGTTTCAACACAAATAG 1677
377 lPheAlaGluAlaPheLeu..... 1683
      ::::: |||
1679 AACTAATAGATTCTATTTTAGTACACTGATATATTTGCTTGCTGCT 1727
384 .....GluHisLeuThrLysLysLeuGluAspProSerAsp 1696
      ::::: |||
1729 AACGACAGCAGACAGAAAGAAATTAAGTATTAACAA..... 1764
384 GAlaIleIleArgGluAlaIleLysAsnTyrIleGlySerPheLeuAla 418
      ::::: |||
1764 .....TATTTGGAGTGTGACATTGCA 1786
418 rAlaLysPheIleProIleGluThrValLysSerCysLeuAspLeuGlu 429
      ::::: |||
1786 GAGCTAAAGAAATTCGACAGACTCAATATTTGCTTCCAGAGTATTTA 1835
418 vAlAsnTyrPleuHisIleTyrLeuAsnAsnGluAspSerGlyThrLysAl 446
      ::::: |||
1835 AGCTGAGGTATGCAAGAGTTTAAACATTCTAGTACAGATTTCACAGAC 1945
446 aPheCysAspValAlaLeuHisGlyProPheTyrSerAlaCysHisVal 463
      ::::: |||
1896 AGTGGAGGTATGCAAGAGTTTAAACATTCTAGTACAGATTTCACAGAC 1945
463 aIlePheTyrThrPheValPheArgHisLys.....GluLeuLeuSerGly 477
      ::::: |||
1936 TTGTTACATTTTGTTCGACATATACATCTTCAGACATACAGATGCT 1985
476 AsnLeuLysGluGlyLeuGluIuhTrLeuGluInsSerLeuAsnPheGluAla 494
      ::::: |||
1985 AACCTGGAGTGTCACTTCGCAAAATTC.....TTCGAAACAT 2024
494 eValMetSerGluLeuAsnProLeuLysIleCysLeuProSerValVal 511
      ::::: |||
2024 GATCATTTCAAAATTTAAACCATTAAGCTTTGTAAGAAATGCAATG 2078
511 snPhePheAlaAlaIleThrAspLysTyrGluLeuValPheCysTyrThr 527
      ::::: |||
2078 TATTTTGTCTGCTATTTCGCAACAAAGAGGCTGCGATGCTTATGCT 2128
528 lIleIleGluAlaAsnAsnArgGluMetLeuProValIleArgSerThrAl 544
      ::::: |||
2128 AATAATGAAATTAAGCAATGAAACATATGACGCGATATCGTGTAAAG 2178
544 aGlyLysAsp.....SerValGluIleCysThrAsn..... 554
      ::::: |||
2178 TATATAGCAGCAAGACAAAGATTCGCTACAGCAAGCAATATCATGCTTT 2228
557 .....ProLeuAspThrPhePhe 560

```



420 ..... 420

154903 TCTTGGTTTCTCTGACCATGGATTATTAAAGAACATATGAA 154952

420 ..... 420

154953 CTACTAGCATGCTTTCTAATGAAAGCAGTGTCTGTTTATGATAT 155002

420 ..... 420

155003 TCTGATTTTATGACGTAAAGTCACTGATTTTCTTTCTTCA 155052

421 ... ThrValIysSerCysLeuAspLeuValAsnTrpLeuHisIleTyr 436

155053 GTACTGTAAACCAATGGCTAGATTTTGTGTTACTGCTGCACATATAC 155102

437 LeuAsnAsnGlnAspSerGlyThrLysAlaPheCysAspValAlaLeuHis 453

155103 CTTAATACCGAGATTCGGACAAAGCAATCTGCAATGTTGCTCTCCA 155152

453 sctIyProPheTyrSerAlaCysGlnAlaValPheTyrThrPheValIlePheA 470

155153 TGGACCATTTACTGACGCTGCGCAAGCTGTGTTCTACACCTTTGTTTAA 155202

470 rHISLysGlnLeuLeuSerGlyAsnLeuLysGluGly 482

155203 GACACAGCACCTTTTAAAGCGAAACCTGAAGAGAGCT 155240

seq\_name: gb\_hlg:AC007615

## seq documentation block:

LOCUS AC007615 175691 bp DNA linear HTG 03-JUL-2001

DEFINITION Homo sapiens chromosome 16 clone RP11-528K16, WORKING DRAFT

SEQUENCE: 1 ordered pieces.

ACCESSION AC007615 GI:14589428

VERSION AC007615.6

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_ACTIVEPIN.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 175691)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

2 (bases 1 to 175691)

Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, F., Robinson, D.,

Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,

Bryant, J., Tesmer, J., Heinicke, L., Longmire, J., White, S., Tatum, O.,

Campbell, C., Fawcett, J., Maithe, M., Bussod, M., Sutherland, R.,

McMurry, K., Han, C. and Beaven, L.

Direct Submission

Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint

Genome Institute, Los Alamos National Laboratory, MS M888, Los

Alamos, NM 87545, USA

On Jul 3, 2001 this sequence version replaced gi:13928651.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

- \* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
- \* This sequence will be replaced
- \* by the finished sequence as soon as it is available and
- \* the accession number will be preserved.

1 175691: contig of 175691 bp in length.

location/qualifiers

1: 175691

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="16"

/clone="RP11-528K16"

/others

BASIS COUNT 42392 a 42125 c 45277 g 45856 t 1 others

ORIGIN

## alignment\_scores:

Quality: 398.00 Length: 346

Ratio: 3.790 Gaps: 1

Percent Similarity: 26.515 Percent Identity: 26.263

## alignment\_block:

US-09-932-678-2 x AC007615

Align seq 1/1 to: AC007615 from: 1 to: 175691

377 GYPhaAlaGluAlaPheLeuGluHisLeuTrpLysLysLeuGlnAspTr 434

3000c GATTCGCCAGACGACATTTTTCGAAACCTTTGCAAAAATTCGAGATTC 40054

353 GserAsnProAlaIleLeuArgGlnAlaAlaGlyAsnTyrIleLysSerP 410

4005c AGTATATCTTCGACATCAICAGCGAGGCTGTGGAAATTTATATGCACT 30104

411 heLeuAlaArgAlaLysPheIleProLeuIle..... 420

30105 TTTTGGCAGAGCTAAATTATATCTCTCTATAGTACGCTAATTTCTT 30154

420 ..... 420

30155 GAGTACCTTTTAAATCATGCTTTAAAGAGCATATGCACTGCTCAAGT 30204

420 ..... 420

420 ..... 420

30205 CAGAAATATCTCCCAATAGTTTTTGGCAAGTTTTTAAAGTAAATAATT 30254

420 ..... 420

420 ..... 420

30255 GCTACTCTGTGCAGATGTTATATTTCTAGTGTGATTTAGAAATAA 30304

420 ..... 420

420 ..... 420

30305 GCTCTGCTTTTCTAATAATAAACCTGATGTAAGGCTTTGTAATCT 30354

420 ..... 420

420 ..... 420

30355 GTTGTGTGCTGTGACTTAAATATGCAATTTGTTGGATATAGAGTA 30404

420 ..... 420

420 ..... 420

30405 GATTGTATATGTTTGAATGCTTTGAGCTTGTGCAATTTTTCATTT 30454

420 ..... 420

420 ..... 420

30455 GATTGCAAAATATATGATATCTTTAAAGCTGAGAAACCGTTTAA 30504

420 ..... 420

420 ..... 420  
40905 GCTAATACACCTATCTAGGAGGCTCAGCTCTCTTCTACGAGCAAGACAT 40954  
420 ..... 420  
40955 ATTCTTGACAAAGGATCTCTGGAGAACTTCCTTCAAGAGAAATTAAT 40604  
420 ..... 420  
40905 CACCAATACACCTACATGAGATGCAACAGCAATATCTCAATCTCCATCAG 40654  
420 ..... 420  
40955 TGTCTTCTTGACAGAGCTTCAAGACATCTCTATTTGAGCAATCAG 40704  
420 ..... 420  
40905 CTTAAAGAGGCTGCTCTTGACATCTCTACAGAACTTCCTTCTATCTCT 40754  
420 ..... 420  
40955 TGGATTTCTTAGAGAGTCAGCTCTCTAGAGAGCTCAGAGCTACTTG 40804  
420 ..... 420  
40905 CTTCTAATCTGAGAGCTTCTCTCTCTACCTACCTCATCTCTATCTCTCT 40854  
420 ..... 420  
40955 TCTTTCTTCTCTCTCTGAGCTTATGCTATTAATTAAGAGAACTATAGAA 40904  
420 ..... 420  
40905 GTTACTTACATACCTTCTAAGGAAAGCTAGCTCTGCTTCTTATCTAT 40954  
420 ..... 420  
40955 TCTTGCTTTTCTTATGAGAGCTAAGCTTGATTTGAATTTCTTCTCTTA 41004  
421 ..... TheValLysSerCysLeuAspLeuValAsnTrpGlnHisIleCys 436  
41005 GAACTCTAAGAACTATGCTTACATCTTTTGCTTAACTGGCTGACATATAC 41054  
437 TGGAGAGAGCTGAGSCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 453  
41055 CTTAATTAATCTAGATTTCTGAGAGCAAGAGATCTCTGCAATCTGCTTCCA 41104  
453 sclytrcphcCysCysAlaCysGlnAlaValIleCysTrpPheValIlePheA 470  
41105 TGGAGCATTTTACTTACCTTGGCTGAGAGCTGCTGCTGCTGCTGCTGCTT 41154  
470 TGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 482  
41155 GAGATAGAGAGCTTCTGAGCTGAGAACTTGAAGAGAGCT 41192















CC responsible for genetic disorders of other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/publicshed\\_pcl\\_sequences](http://wipo.int/pub/publicshed_pcl_sequences).

XX Sequence 2410 BP; 672 A; 437 C; 547 G; 754 T; 0 other;

Query Match 40.98; Score 639.4; DB 23; Length 2410;  
 Best Local Similarity 96.88; Pred. No. 1.4e-177;  
 Matches 706; Conservative 0; Mismatches 16; Indels 7; Gaps 5;

QY 1347 gatcttcttggttaactgagcagatataaccataaccagatctggaacaaagca 1406  
 DB 1 gatcttcttggttaactgagcagatataaccataaccagatctggaacaaagca 60  
 QY 1407 tctatgagatgtgtcttcacatgagacatttactcagcctggaacatctatcavacc 1466  
 DB 61 tctcagagatgtgtcttcacatgagacatttactcagcctggaacatctatcavacc 120  
 QY 1467 ttgttttttagacacaaagcagctttagcgcgaaacctgaagaagtttgcagatctt 1526  
 DB 121 ttgttttttagacacaaagcagctttagcgcgaaacctgaagaagtttgcagatctt 180  
 QY 1527 cagatctgaattttagacgagatagatagcagcagatlaalccctcgaagat-ttgcct 1585  
 DB 181 cagatctgaattttagagtagatagatagcagcagatlaalccctcgaagatagagctt 240  
 QY 1586 gcccctcagtggttaacttttctcgcacatccacaataaataccagctcctcttgcga 1645  
 DB 241 gcccctcagtggttaacttttctcgcacatccacaataaataccagctcctcttgcga 300  
 QY 1646 cccctcctttagagagacacatccagatgctgcagctatctagagtagcgcgtgcag 1705  
 DB 301 cccctcctttagagagacacatccagatgctgcagctatctagagtagcgcgtgcag 360  
 QY 1706 aaacatcagtgagatctcacaacaccgcagagacattcttcccttt-gatcctctga 1764  
 DB 361 aaacatcagtgagacatccacaacaccgcagagacattcttccctttgagctctgag 420  
 QY 1765 tctcgaagagtgcaagaatt-cattgatcctattt-attcaggtatggaagacatg 1820  
 DB 421 tctcgaagagtgcaagaattcattgatcctatttattcaggtatggaagacatg 480  
 QY 1821 agtctgaagagtgcaagaattcacaagaaccatgagaagacatagtgagaagaa 1880  
 DB 481 agtctgaagagtgcaagaattcacaagaaccatgagaagacatagtgagaagaa 540  
 QY 1881 gattgatgattctcgaagagcgaagtgcaccagaatlaalacgttgattgagatcacaca 1940  
 DB 541 gattgatgattctcgaagagcgaagtgcaccagaatlaalacgttgattgagatcacaca 600  
 QY 1941 agtctccttctgaaagcattcccgagatcctcctcgaagtatgagctcccccacatgtt 2000  
 DB 601 agtctccttctgaaagcattcccgagatcctcctcgaagtatgagctcccccacatgtt 660  
 QY 2001 taca-tgacaaccagatccctcctcgaagcagaagaattgttactgagatctgacatttgc 2059  
 DB 661 taca-tgacaaccagatccctcctcgaagcagaagaattgttactgagatctgacatttgc 720  
 QY 2060 attccccc 2068  
 DB 721 attccccc 729

RESULT 6  
 AAS94253  
 ID AAS94253 standard: cDNA; 776 BP.  
 XX  
 AC AAS94253;

XX 13-EPB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #20057.  
 DE  
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 PN W6200175067-A2.  
 XX  
 PD 11-MAY-2001.  
 XX  
 PR 30-MAR-2001; 2001W6-0508631.  
 XX  
 PR 31-MAR-2000; 2000J5-0546217.  
 XX  
 PR 23-MAY-2000; 2000J5-0549167.  
 XX  
 PA (HVSF-) HVSFQ IM-1.  
 XX  
 PT Dramatic RT, Liu C, Tang YT.  
 XX  
 DR WPI: 2001-639362/73.  
 DR P-IDS; AHC28056.  
 XX  
 PT New isolated polypeptide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 1; SEQ ID NO 28057; 103bp; English.  
 XX  
 CC The invention relates to isolated polypeptide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polypeptides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity or (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC treatment of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity  
 CC The polypeptide and polynucleotide sequences have applications in:  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/publicshed\\_pcl\\_sequences](http://wipo.int/pub/publicshed_pcl_sequences).

XX Sequence 776 BP; 205 A; 157 C; 198 G; 215 T; 0 other;

Query Match 27.58; Score 568.6; DB 23; Length 776;  
 Best Local Similarity 97.88; Pred. No. 5.6e-157;  
 Matches 629; Conservative 0; Mismatches 9; Indels 5; Gaps 5;

QY 4 gaaactgtgacttgaagaagatgagatcgaagctgaagcgaagcagatcgaattatctg 63  
 DB 133 gaaactgtgacttgaagaagatgagatcgaagctgaagcgaagcagatcgaattatctg 192  
 QY 4 gaaactgtgacttgaagaagatgagatcgaagctgaagcgaagcagatcgaattatctg 122  
 DB 133 gaaactgtgacttgaagaagatgagatcgaagctgaagcgaagcagatcgaattatctg 252  
 QY 124 tctgcaatcaagaacatgtgctgctcgaagatgagatcttcaaatatgcatatcaga 181  
 DB 253 tctgcaatcaagaacatgtgctgctcgaagatgagatcttcaaatatgcatatcaga 412







RESULT 10  
 ABA83018  
 ID ABA83038 standard: DNA: 1461 bp.  
 AC  
 XX ABA83038:  
 XX  
 XX 05-FEB-2002 (first entry)  
 XX  
 DE Human transcription factor TRFX-65 coding sequence.  
 DE  
 XX Human transcription factor: TRFX; cell proliferative disease;  
 KW autoimmune disease; inflammation; neurological disease;  
 KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;  
 KW neuroprotective; antiinflammatory; gene therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 XX W6200172777-A2.  
 XX  
 XX 04-OCT-2001.  
 XX  
 XX 13-MAR-2001: 2001W0-US08117.  
 XX  
 XX 13-MAR-2000: 2000US-0188986.  
 XX  
 XX (INCYTE GENOMICS INC.  
 XX  
 XX Hillman JL, Hauduin MR, Yue H, Lal P, Lu DAM, Patterson C;  
 P1 Azimzal Y, Bandman O, Tani YT, Mathur P, Shah P, Au-Young J;  
 P1 Reddy R;  
 XX  
 XX WPI: 2001-570886/64.  
 DR P-FSDB: ABR50214.  
 XX  
 XX Novel transcription factor polypeptides, used to treat diseases  
 PT associated with altered activity and expression of TRFX, and to screen  
 PT for agents capable of modulating its activity -  
 XX  
 XX  
 XX Claim 11: Page 299; 327pp; English.  
 CC The present sequence is the coding sequence for a human transcription  
 CC factor. The transcription factor and its coding sequence are useful in  
 CC the diagnosis, treatment and prevention of diseases associated with  
 CC altered expression of the transcription factor e.g. cell proliferative,  
 CC autoimmune/inflammatory, neurological and developmental disorders. A  
 CC number of specific disorders/diseases are given in the specification,  
 CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
 CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
 CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
 CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
 CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
 CC disease, stroke, and viral, bacterial, fungal and protozoal infections.  
 XX  
 XX Sequence 1461 bp; 413 A; 335 C; 298 G; 415 T; 0 other;  
 SO  
 Query Match 20.3%; Score 420.8; DB 22; length 1461;  
 Best Local Similarity 97.3%; Pred. No. 31e-114;  
 Matches 428; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 QY 1188 tttcaattgaattcgaagaagcatttggaaacatctcgaanaaatcgaagaccca 1247  
 Db 530 tttcaatattgattcgaagaagcatttggaaacatctcgaanaaatcgaagaccca 589  
 QY 1248 attaatcctgcatcattcgaagcctctcgaagaattatattatattggaagc 1307  
 Db 590 attaatcctgcatcattcgaagcctctcgaagaattatattatattggaagc 649  
 QY 1408 gctcaaatatttctcttatttataataaattcgaagcatttgaatttgaattgagc 1367  
 Db 650 gctcaaatatttctcttatttataataaattcgaagcatttgaatttgaattgagc 709

QY 1468 gacatataccttaataaagagatttcgaagaagaagcattcgaagatttgcctccat 1427  
 Db 710 gacatataccttaataaagagatttcgaagaagaagcattcgaagatttgcctccat 769  
 QY 1428 ggaacatttactcagccttcgaagcctgattctcagccttattttacacacacagc 1487  
 Db 770 ggaacatttactcagccttcgaagcctgattctcagccttattttacacacacagc 829  
 QY 1468 ctcttaagcagaacacccaagaagatttcgaagattctcgaagccttattttacac 1547  
 Db 830 ctcttaagcagaacacccaagaagatttcgaagattctcgaagccttattttacac 889  
 QY 1518 atagagatagacagcctgaatcctcctgaagaatttactcagccttgaatttttt 1507  
 Db 650 atagagatagacagcctgaatcctcctgaagaatttactcagccttgaatttttt 949  
 QY 1608 actgaatcagaataaata 1627  
 Db 950 actgaatcagaataaata 969  
 RESULT 11  
 AAX51662  
 ID AAX51663 standard: cDNA: 437 bp.  
 AC  
 XX AAX51663:  
 XX  
 XX 21-JUN-1999 (first entry)  
 XX  
 XX Human secreted protein 5' EST SFY 10 M8242.  
 DE  
 XX  
 XX Human secreted protein: EST; expressed sequence tag; diagnosis;  
 KW for cancer; gene therapy; chromosome mapping; signal peptide;  
 KW TGF-beta1; regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemokine; chemokine; haematopoietic  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
 XX  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX W60906549-A2.  
 XX  
 XX 11-FEB-1999.  
 PD  
 XX 31-JUL-1998: 98WD-1B01231.  
 PF  
 XX 01-AUG-1997: 97US-0905279.  
 PR  
 XX (GIST) GENSET.  
 XX  
 XX Duclert A, Dubas ML, Edwards J, Lacroix B;  
 P1  
 XX WPI: 1999-153779/13.  
 DR P-FSDB: AAY12885.  
 XX  
 XX New nucleic acids encoding human secreted proteins obtained from  
 PT cDNA libraries derived from testis, ovary, uterus and spleen tissue  
 PT  
 XX (claim 1: Page 344-345; 522pp; English.  
 PS  
 XX AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12481 to  
 CC AAY12493, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, haematopoiesis regulation  
 CC activity, tissue growth regulation activity, reproductive hormone  
 CC regulating activity, chemokine/chemokine activity, haematopoietic  
 CC and angiogenic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in for example, gene therapy and chromosome mapping procedures.









PR 31-MAR-2000: 2000US-0540217.  
 PR 23-AUG-2000: 2000US-0649167.

XX (HYSE-) HYSEO INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: AR623449.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 1: SEQ ID No 2340: 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPo  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 1017 BP: 301 A: 214 C: 230 G: 272 T: 0 other:

Query Match 15.2%; Score 315; DB 23; Length 1017;

Best local Similarity 91.7%; Pred. No. 46-82;

Matches 333; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1188 ttcgaattgaattcgaagagcatlltggaaacatctcggaaaattgcaggaccga 1247  
 || || |||||  
 DB 445 tttaaatgagatcgcaagacatlltggaaacatctcggaaaattgcaggaccga 504  
 QY 1248 agtaatccctgacatcagcagcagctctggaatlatattgaagctlltggcaaga 1307  
 |||||  
 DB 505 aglaalccctgacatcagcagcagctctggaatlatattgaagctlltggcaaga 564  
 QY 1308 gctaaatttctcttcttattcgtgaataatcgtcagatcctlltggtaactgacta 1367  
 |||||  
 DB 565 gctaaatttctcttattcgtgaataatcgtcagatcctlltggtaactgacta 624  
 QY 1368 caacataccttaataaccagattcggaaacaagcattctgaatgttgccttccaa 1427  
 |||||  
 DB 625 caacataccttaataaccagattcggaaacaagcattctgaatgttgccttccaa 684  
 QY 1428 gaacatttlaactcaagccttgcgaactgtgtctacacctllgtlltlaacacaagaag 1487  
 |||||  
 DB 685 gaacatttlaactcaagccttgcgaactgtgtctacacctllgtlltlaacacaagaag 744  
 QY 1488 cttttgagcgaagaacctggaagaagaatttgcagatctctgaactctgaatttgaagcag 1547  
 |||||  
 DB 745 cttttgagcgaagaacctggaagaagaatttgcagatctctgaactctgaatttgaagcag 804  
 QY 1548 ata 1550  
 ||  
 DB 805 aaa 807

Search completed: July 30, 2002, 03:59:40  
 Job time: 4839 sec

...